Australian Government

Department of Defence

Defence Science and
Technology Organisation

A Course of Lectures on Statistical Mechanics

Don Koks

Electronic Warfare and Radar Division

Defence Science and Technology Organisation

DSTO–GD–0612

ABSTRACT

This is a set of lectures given by the author in 2009 at Flinders University, Adelaide, comprising one semester of a third-year undergraduate course in physics. The lectures begin with an introduction to the theoretical background of statistical mechanics, and then continue with a mixture of theory and application. Topics covered are those that comprise the standard tool kit for advanced study in the field.

APPROVED FOR PUBLIC RELEASE

This report is an amended version of the original published in June 2010.
Three typographical mistakes have been fixed: on page 33, "0" has been changed
to "negligible" in equation (9.5); on page 52 "0" has been changed to "−∞" in the first line
of (12.4); and on page 73 an incorrect integral in (16.39) has been removed.

APPROVED FOR PUBLIC RELEASE

A Course of Lectures on Statistical Mechanics
Executive Summary

These lectures were given by the author in 2009 for a one-semester course at Flinders University, Adelaide, as part of that university's third-year undergraduate course in physics.

The lecture notes begin with an introduction to the mathematical background of statistical mechanics. They introduce the all-important notion of entropy, which leads to the concept of temperature and then to the basics of thermodynamics. Following this is an excursion into the physical chemistry of dissolved salts, and the idea of a reaction attaining equilibrium. The Boltzmann distribution is then introduced; this serves as the departure point for a study of systems interacting with an environment. The relevant ideas allow the entropy of more complex and non-isolated systems to be calculated. At this point we include a discussion of the approach to entropy advocated by E.T. Jaynes, who was at the forefront of advanced ideas in statistical mechanics throughout the twentieth century.

Several standard topics are then covered: the Maxwell speed and velocity distributions, and the theory of transport processes that successfully interrelates thermal conductivity, viscosity, and heat capacity; this success was historically of prime importance to the formation of an atomic view of matter in physics.

The notes end with a discussion of quantum statistics, blackbody radiation, electric conductivity, and semiconductors.

It should be emphasised that, course notes being what they are, the following pages cover the main ideas briskly. In particular, no pictures have been included, although of course many were drawn in the lectures to aid in the presentation.

Author

Don Koks
Electronic Warfare and Radar Division

Don Koks completed a doctorate in mathematical physics at Adelaide University in 1996, with a thesis describing the use of quantum statistical methods to analyse decoherence, entropy and thermal radiance in both the early universe and black hole theory. He holds a Bachelor of Science from the University of Auckland in pure and applied mathematics and physics, and a Master of Science in physics from the same university with a thesis in applied accelerator physics (proton-induced X ray and γ ray emission for trace element analysis). He has worked on the accelerator mass spectrometry programme at the Australian National University in Canberra, and in commercial Internet development.

Currently he is a Research Scientist with the Maritime Electronic Warfare Systems group in the Electronic Warfare and Radar Division at DSTO, specialising in geospatial orientation concepts. He is the author of the book *Explorations in Mathematical Physics: the Concepts Behind an Elegant Language* (Springer, 2006).

Contents

1   Introduction

These lectures begin with an introduction to the mathematical background of statistical mechanics. They introduce the all-important notion of the entropy of an isolated system. This leads to the concept of temperature and then to the basics of thermodynamics. Following this is an excursion into chemical concepts to do with dissolved salts and the idea of a reaction attaining equilibrium. Next, the Boltzmann Distribution is introduced; this serves as the departure point for the study of systems interacting with an environment of which we might know nothing. The ideas here give meaning to the entropy of more complex and non-isolated systems, and allow it to be calculated. At this point we include a discussion of the ideas of E.T. Jaynes, who was at the forefront of advanced ideas in statistical mechanics throughout the twentieth century.

Following this, several standard topics are covered: the Maxwell speed and velocity distributions, and the theory of transport processes that successfully interrelates thermal conductivity, viscosity, and heat capacity; this success was historically of prime importance in the formation of an atomic view of matter in physics.

The notes end with a discussion of quantum statistics, blackbody radiation, electric conductivity, and semiconductors.

The general ordering of the subjects here follows reference [1]. However, much of the mathematical analysis in these notes follows different routes from those in that book, and other content has been added to these notes.

It should be emphasised that, course notes being what they are, the following pages do cover the main ideas briskly. In particular, no pictures have been included, although of course many were drawn in the lectures to aid in the presentation.

2   Preliminaries for Counting Large Numbers

Statistical mechanics is built on the idea that the world can be described using probability. Yet on the surface there seems to be little randomness in the world around us, so is a probabilistic description really such a good idea?

We will begin to answer this question by asking something more basic: given a set number of particles of a gas in a room, what's the chance of there being some given number of particles in a given part of the room? Furthermore, how probable are *fluctuations* around this number? It will turn out that for systems with large numbers of particles such as we find in everyday life, fluctuations are very improbable things. This indicates that a probabilistic view of the world might well be compatible with the fact that we don't see a lot of randomness around us. It's the starting point for the subject of statistical mechanics.

2.1   The Binomial Distribution

It's simplest to divide the room into two regions and find probabilities for different numbers of the particles to be in each region. This is the job of the *binomial distribution*. Given N distinguishable particles, allocate each to one of two bins. The chance of a particular particle being allocated to bin 1 is p, so that the chance of a particular particle being

allocated to bin 2 must be $1 - p$. What is the chance $P(n)$ that we'll find any n particles (without regard for order) in bin 1?

The chance that any such combination occurs, with n particles in bin 1 and $N - n$ particles in bin 2, is $p^n (1 - p)^{N-n}$. We need only count how many such combinations there can be. Do this by labelling the particles $1, 2, \ldots, N$ and simply writing down all possible combinations. We can do this systematically by writing down all *permutations* as if the particles were all lined up in a row. This keeps track of their order, which allows us to count them more easily since now there are simply $N!$ possible permutations. For the case of $N = 7$ particles in total, $n = 3$ of which appear in bin 1, we might write all 7! permutations as (with bin 1 written first, then a space, then bin 2)

$$
7! \text{ rows}
\begin{cases}
\left.
\begin{matrix}
1\,2\,3 & 4\,5\,6\,7 \\
1\,2\,3 & 4\,5\,7\,6 \\
& \vdots \\
1\,3\,2 & 4\,5\,6\,7 \\
1\,3\,2 & 4\,5\,7\,6 \\
& \vdots
\end{matrix}
\right\} 3!\,4! \text{ rows} \\[2em]
\left.
\begin{matrix}
1\,2\,4 & 3\,5\,6\,7 \\
1\,2\,4 & 3\,5\,7\,6 \\
& \vdots
\end{matrix}
\right\} 3!\,4! \text{ rows} \\[1em]
\qquad\quad \text{etc.}
\end{cases}
\tag{2.1}
$$

Each combination appears 3! 4! times, so the total number of *permutations*, 7!, over-counts the number of *combinations* by this factor. Hence the number of combinations is $7!/(3!\,4!)$.

> Alternatively, we could focus on bin 1 and note that there are $7 \times 6 \times 5 = 7!/4!$ ways of putting three particles into it if we take order into account (i.e. permutations); to count combinations, we must correct for the fact that each combination produces 3! permutations, so must divide the number of permutations by 3! to get $7!/(3!\,4!)$. You might like to ponder on how to extend this approach to the case of many bins that we'll examine in Section 11.

More generally, the total number of combinations is $N!/[n!\,(N - n)!]$, also written ${}^{N}C_n$. Each of these combinations occurs with probability $p^n (1 - p)^{N-n}$, so the final sought-after probability is

$$
P(n) = \frac{N!}{n!\,(N - n)!}\, p^n (1 - p)^{N-n}.
\tag{2.2}
$$

This function of the number of particles n is called the binomial distribution.

Example 1: 5 molecules are in a room. What's the chance that any 2 of them are in the front $1/3$ at some chosen moment?

$$
\underbrace{2 \text{ molecules}}_{\text{front } 1/3} \quad \underbrace{3 \text{ molecules}}_{\text{back } 2/3}
$$

$$
\text{Prob.} = \frac{5!}{2!\,3!}\, (1/3)^2 (2/3)^3 \simeq 0.33. \qquad \underline{\underline{\text{Answer}}}
\tag{2.3}
$$

Example 2: 10 molecules are in a room. What's the chance that any 4 of them are in the front $1/3$ at some chosen moment?

$$
\underbrace{4 \text{ molecules}}_{\text{front } 1/3} \quad \underbrace{6 \text{ molecules}}_{\text{back } 2/3}
$$

$$\text{Prob.} = \frac{10!}{4!\,6!}\,(1/3)^4\,(2/3)^6 \simeq 0.23\,. \qquad \underline{\text{Answer}} \qquad (2.4)$$

For larger factorials, use *Stirling's rule*:

$$n! \sim n^{n+1/2}\,e^{-n}\,\sqrt{2\pi}\,. \qquad (2.5)$$

By "$\sim$" we mean LHS/RHS $\to 1$ as $n \to \infty$, although their difference might not go to zero. In particular,

$$\ln n! \sim (n + 1/2)\ln n - n + \ln\sqrt{2\pi} + \frac{1}{12n} - \frac{1}{360n^3} + \ldots \qquad (2.6)$$

This is an example of an *asymptotic series*. To see how it differs from the more well known *convergent* series, consider how a convergent series is used: if (2.6) were such a series, we could fix n and ensure convergence by letting the number of terms go to infinity. In other words, we could calculate $\ln n!$ to any accuracy by summing a sufficient number of terms.

In contrast, and somewhat bizarrely, an asymptotic series does not converge in this way for any value of n. The coefficients of the first few powers of n in (2.6) start out by decreasing term by term, but that trend soon reverses and they become very large. For any choice of n, they soon grow larger at a faster rate than the (denominator) powers of n, so that the series can never converge. To use (2.6), we must truncate its right-hand side after some arbitrary number of terms, and then note that increasing n gives a different sort of convergence: $\lim_{n\to\infty} \frac{\text{LHS}}{\text{truncated RHS}} = 1$. That means we can't use the series to calculate $\ln n!$ to arbitrary accuracy. Precisely where the truncation might best be made to maximise the accuracy of the approximation is something of an art.

Example 3: 1000 molecules are in a room. What's the chance that any 400 of them are in the front $1/3$ at some chosen moment?

400 molecules	600 molecules
front $1/3$	back $2/3$

$$\text{Prob.} = \frac{1000!}{400!\,600!}\,(1/3)^{400}\,(2/3)^{600}\,. \qquad (2.7)$$

$$\begin{aligned}
\text{Then } \ln\text{prob.} \simeq\ & 1000.5\ln 1000 - \cancel{1000} + \cancel{\ln\sqrt{2\pi}} \\
& - 400.5\ln 400 + \cancel{400} - \cancel{\ln\sqrt{2\pi}} \\
& - 600.5\ln 600 + \cancel{600} - \ln\sqrt{2\pi} \\
& + 400\ln 1/3 + 600\ln 2/3 \simeq -13.3716\,,
\end{aligned}$$

$$\text{so prob.} \simeq 1.559 \times 10^{-6}. \qquad \underline{\text{Answer}} \qquad (2.8)$$

(A presumably fairly exact answer from Mathematica is $\simeq 1.558 \times 10^{-6}$.)

N.B. Some books write $\ln n! \simeq n\ln n - n$. This will give $\ln\text{prob.} \simeq -9.7$ in Example 3, so is clearly not accurate in this case—and it becomes more and more inaccurate as $n \to \infty$. But compare this expression with the correct one when $n \approx 10^{24}$, and ask yourself whether it might in fact be useful after all.

2.2 Quantifying Fluctuations in the Binomial Distribution

Consider $N = 10$ coins flipped. What is the chance that half land heads up?

$$P(5 \text{ heads}) = \frac{10!}{5! \, 5!} \, (1/2)^5 \, (1/2)^5 \simeq 0.25 \,. \tag{2.9}$$

Now for $N = 100$:

$$P(50 \text{ heads}) = \frac{100!}{50! \, 50!} \, (1/2)^{50} \, (1/2)^{50} \simeq 0.08 \,. \tag{2.10}$$

And now $N = 10^6$:

$$P(500{,}000 \text{ heads}) = \frac{1{,}000{,}000!}{500{,}000! \; 500{,}000!} \, (1/2)^{500{,}000} \, (1/2)^{500{,}000} \simeq 0.0008 \,. \tag{2.11}$$

Although the chance of getting n heads is maximal if $n = N/2$, it goes to 0 as $N \to \infty$. A more useful question: given N, p, what is the value of n where $P(n \text{ heads})$ peaks, and what is a good measure of the width of the probability distribution? Or, rather than ask where the peak lies, ask: what is the mean number of heads , often written $\bar{n}$ or $\langle n \rangle$?

Basic probability theory gives $\bar{n} = pN$. Alternatively we can use a first-principles approach to write

$$\bar{n} = \sum_{n=0}^{N} nP(n) = \sum_{n=0}^{N} n \; {}^N C_n \, p^n \, (1-p)^{N-n} \,. \tag{2.12}$$

This looks to be a difficult expression to evaluate. But we can do it using a kind of trick: replace $1 - p$ (when it appears explicitly) by "q" and treat q initially as an independent variable, only setting it equal to $1 - p$ at the end of the calculation. Now make use of two expressions:

$$(p \, \partial_p)^k \, p^n = n^k \, p^n \quad \text{and} \quad (p + q)^N = \sum_{n=0}^{N} {}^N C_n \, p^n q^{N-n} \,, \tag{2.13}$$

where $\partial_p \equiv \partial/\partial p$. Use the first of these in (2.12) with $k = 1$, then the second, to write

$$\bar{n} = \sum_n {}^N C_n \, p \, \partial_p \, p^n \, q^{N-n} = p \, \partial_p (p + q)^N = pN(p + q)^{N-1} = pN \,, \tag{2.14}$$

as expected. For the measure of fluctuation, use the variance $\sigma^2 = \overline{n^2} - \bar{n}^2$. Again use (2.13)—now with $k = 2$—to write

$$\overline{n^2} = \sum_n n^2 P(n) = \sum_n n^2 \; {}^N C_n \, p^n \, q^{N-n} = (p \, \partial_p)^2 \sum_n {}^N C_n \, p^n \, q^{N-n}$$

$$= \cdots = p^2 N^2 + Np(1-p) \,. \tag{2.15}$$

So $\sigma^2 = Np(1-p)$. Define the *relative fluctuation* $\equiv \sigma/\bar{n} \propto 1/\sqrt{N}$.

Example 1. 10 molecules in a room. What is the mean number in the front third of the room, and what is its relative fluctuation?

$$\bar{n} = pN = 1/3 \times 10 = 3\,1/3 \,. \quad \underline{\text{Answer}}$$

$$\sigma = \sqrt{Np(1-p)} = \sqrt{20}/3, \text{ so } \sigma/\bar{n} \simeq 0.45 \,. \qquad \underline{\text{Answer}} \qquad (2.16)$$

Example 2. Do the same for 10^{27} molecules in a room—a realistic figure.

$$\bar{n} = pN = 1/3 \times 10^{27} \,. \qquad \underline{\text{Answer}}$$

$$\sigma/\bar{n} = \sqrt{(1-p)/\bar{n}} = \sqrt{\frac{2/3}{1/3 \times 10^{27}}} \simeq 4.5 \times 10^{-14} \,. \qquad \underline{\text{Answer}} \qquad (2.17)$$

This is tiny! Systems with stupendously large numbers of entities are *very* predictable.

2.3 The Gaussian Approximation to the Binomial Distribution

Calculating $P(n) = {}^{N}C_n \, p^n \, (1-p)^{N-n}$ via Stirling is tedious and doesn't give a feel for $P(n)$. Let's do better by approximating its logarithm by a Taylor Series. (Why its logarithm? Because this is less peaked and so yields a better approximation.)

$$f(n) \equiv \ln P(n) = f(\bar{n}) + f'(\bar{n})(n - \bar{n}) + \frac{1}{2!}f''(\bar{n})(n - \bar{n})^2 + \ldots \qquad (2.18)$$

$$\begin{aligned} f(n) &\simeq \ln N! - \ln n! - \ln(N-n)! + n \ln p + (N-n) \ln q \\ &\simeq (N + 1/2) \ln N - (n + 1/2) \ln n - (N - n + 1/2) \ln(N - n) \\ &\quad - \ln \sqrt{2\pi} + n \ln p + (N-n) \ln q \,. \end{aligned} \qquad (2.19)$$

Differentiating,

$$f'(n) \simeq -\ln n - \frac{1}{2n} + \ln(N-n) + \frac{1}{2(N-n)} + \ln p - \ln q \,,$$

$$f''(n) \simeq \frac{-1}{n} + \frac{1}{2n^2} - \frac{1}{N-n} + \frac{1}{2(N-n)^2} \,. \qquad (2.20)$$

Thus, if p isn't close to 0 or 1 (which are statistically uninteresting cases anyway),

$$f(\bar{n}) \simeq -\ln \sqrt{2\pi\sigma^2} \,, \quad f'(\bar{n}) \simeq \frac{2p-1}{2\sigma^2} \,, \quad f''(\bar{n}) \simeq \frac{-1}{\sigma^2} \,. \qquad (2.21)$$

So, with $x \equiv n - \bar{n}$,

$$f(n) \simeq -\ln \sqrt{2\pi\sigma^2} - \frac{x^2 - (2p-1)x}{2\sigma^2} \,. \qquad (2.22)$$

Complete the square to give

$$P(n) \propto \exp \frac{-\left[n - (\bar{n} + p - 1/2)\right]^2}{2\sigma^2} \,. \qquad (2.23)$$

Remember that $\bar{n} \gg 1$, so this is a gaussian centred around $\bar{n}$ approximately, with width σ. It's usual to approximate it by

$$P(n) \simeq \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(n - \bar{n})^2}{2\sigma^2} \,, \quad \begin{cases} \bar{n} = Np \\ \sigma^2 = Np(1-p) \end{cases} \qquad (2.24)$$

Example: With $N = 10^{27}$ molecules in a room, what's the chance that the mean number in the front third of the room actually occupies that front third?

N is so large here that we should use (2.24). The mean number in the front third of the room is $\bar{n} = 1/3 \times 10^{27}$. Set $n = \bar{n}$ in (2.24) to get

$$\sigma = \sqrt{20}/3 \times 10^{13}, \quad P(\bar{n}) \simeq \frac{1}{\sqrt{20}/3 \times 10^{13} \times \sqrt{2\pi}} \simeq 3 \times 10^{-14}. \qquad \underline{\underline{\text{Answer}}} \quad (2.25)$$

What's the chance that this number fluctuates upwards by 1%?

$$P(1.01\bar{n}) \simeq P(\bar{n}) \exp \frac{-(0.01\bar{n})^2}{2\sigma^2} \simeq P(\bar{n}) \exp\left(1/4 \times 10^{23}\right)$$

$$\approx 10^{-10^{22}} \qquad \underline{\text{Answer}} \qquad (2.26)$$

This is *very* small: even a 1% fluctuation can be treated as never occurring. More realistically, what's the chance that the occupation number fluctuates by *at least* 1% up or down? With more effort we can show that the answer is "close" to the number in (2.26), but we won't do that calculation here.

2.4   Integrating a Gaussian Function

Gaussian functions are common throughout probability theory, and statistical mechanics is no exception. You will often find yourself integrating them, so here is a good place to write down a general expression for the gaussian integral in terms of the *error function* erf x. First in one dimension,

$$\boxed{\int e^{-ax^2 + bx}\, \mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, e^{\frac{b^2}{4a}}\, \mathrm{erf}\left(\sqrt{a}\, x - \frac{b}{2\sqrt{a}}\right).} \qquad (2.27)$$

This expression is true for all values of a and b—even complex ones. To help visualise it, note that erf x is a strictly increasing odd function over the reals. It's shaped much like $\tan^{-1} x$ for real x, except that erf $\infty = 1$. A special case of (2.27) is

$$\int_{-\infty}^{\infty} e^{-ax^2 + bx}\, \mathrm{d}x = \sqrt{\frac{\pi}{a}}\, e^{\frac{b^2}{4a}}. \qquad (2.28)$$

For a complex integration, it's useful to remember that $\mathrm{erf}(-z) = -\mathrm{erf}\, z$ for all complex z, and erf $z \to 1$ as $|z| \to \infty$, provided $|\arg z| < \pi/4$.

The corresponding definite integral in multi dimensions also comes in handy. Suppose the n integration variables $x_1, \ldots, x_n$ are written as a column vector $\boldsymbol{x}$, A is a real symmetric $n \times n$ matrix, and $\boldsymbol{b}$ is a column vector. Then, with "t" denoting transpose,

$$\int_{-\infty}^{\infty} \exp\left(-\boldsymbol{x}^t A \boldsymbol{x} + \boldsymbol{b}^t \boldsymbol{x}\right)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n = \frac{\pi^{n/2} \exp\left(\boldsymbol{b}^t A^{-1} \boldsymbol{b}/4\right)}{\sqrt{\det A}}. \qquad (2.29)$$

3 Accessible States and the Fundamental Postulate of Statistical Mechanics

At the heart of statistical mechanics is the idea of counting the number of states that a system can occupy. An isolated system is in *equilibrium* when the probabilities that it will be in various states are constant over time. The characteristic time needed for a perturbed system to attain equilibrium is called its *relaxation time*. In this course we'll always assume that systems are always at, or arbitrarily close to, equilibrium. This requires that all changes happen slowly compared to the relaxation time. Such processes are called *quasistatic*.

The fundamental postulate of statistical mechanics is

> An isolated system in equilibrium is equally likely to be in any state accessible to it.

The total number of states accessible at some energy is called Ω. So the chance of the system being found in any particular state is $1/\Omega$.

Example 1: 3 coins are flipped. What is the chance that 2 of them land heads up?

$$P(hh) = {}^3C_2 \, (1/2)^2 \, (1/2)^1 = 3/8 \,. \qquad \underline{\text{Answer}} \tag{3.1}$$

The total number of states of 3 flipped coins is $\Omega = 8$. The number of states with 2 heads is called the *degeneracy* of the 2-heads state, which is 3 in this case.

Example 2: 3 identical particles, spin $1/2$. What is the chance that 2 particles have spin up? Now $\Omega = 4$, corresponding to all down, 1 up, 2 up, 3 up. There's no degeneracy—or equivalently, we might choose to say that the degeneracy equals 1, so

$$P(2 \uparrow) = 1/4 \,. \qquad \underline{\text{Answer}} \tag{3.2}$$

3.1 Density of States for a Monatomic Gas

What is the number of states Ω for a gas of N distinguishable particles of the same mass? The number of states at any particular energy E is generally extremely difficult, if not impossible, to compute. But we can use the fact that the energy spacing between neighbouring states is typically so tremendously small that the state energies can be treated as a continuum. This is analogous to treating the mass of a ruler as distributed continuously along its length. The mass is not really a continuum; it's located in the nuclei of the atoms that comprise the ruler. We cannot talk about the mass at a point a distance L from one end. For this reason, the concept of mass *density* was invented: the mass density is $\varrho(L)$ at a point a distance L from one end, and we calculate it by averaging over many nuclei. We can use this density to calculate approximately how much mass is in some small length ΔL of the ruler: it is $\Delta M(L) \simeq \varrho(L) \, \Delta L$. We even refer to an infinitesimal mass $dM = \varrho(L) \, dL$ even though, strictly speaking, this has no proper physical meaning for a ruler made of atoms.

Since discussing the number of states $\Omega(E)$ at some energy E is often problematic, we treat $\Omega(E)$ like the mass of the ruler at some point. That is, just as we modelled the

ruler as a continuum of mass and worked only with the mass *density* at a point, likewise we approximate the spread of states as a continuum, and define a *density of states*. To do this, write the total number of states in the energy range $0 \to E$ as $\Omega_{\text{tot}}(E)$, like the total mass of a ruler of length L. With a suitable "coarse graining", $\Omega_{\text{tot}}(E)$ is related to the density of states $g(E)$ by $\mathrm{d}\Omega_{\text{tot}} = g(E)\,\mathrm{d}E$. The number of states in a small energy interval will then be $g(E)$ times the width of the interval. If the typical spacing of energy levels around E is ΔE (in practice an incredibly tiny number), then we might write the number of states "at" E as

$$\Omega(E) \simeq \Delta\Omega_{\text{tot}} \simeq g(E)\,\Delta E. \tag{3.3}$$

Hence we can calculate $g(E)$ by first finding $\Omega_{\text{tot}}(E)$ and then writing

$$g(E) = \frac{\mathrm{d}\Omega_{\text{tot}}}{\mathrm{d}E}. \tag{3.4}$$

So rather than try to calculate $\Omega(E)$ for a gas of N distinguishable particles, we'll instead calculate $g(E)$ via $\Omega_{\text{tot}}(E)$. First, consider the number of states accessible to one particle of energy E and momentum $\boldsymbol{p}$. In *phase space* (position–momentum space), the particle has a range $[x]$ available to it in the x direction (similarly $[y], [z]$), and a range $[p_x]$ available to it in the p_x direction (similarly $[p_y], [p_z]$). Quantum mechanically, the particle's position and momentum (in the x direction) are defined only up to $[x], [p_x]$ with at best $[x][p_x] \approx h$, so we partition the phase space into cells where each cell defines one accessible state. For example, for one-dimensional motion the xp-space is divided into cells of area $[x][p_x] = h$.

> Does it make sense to define the volume of one cell of phase space by dividing by one factor of h for each dimension? Perhaps we should use $\hbar/2$ instead? In fact, it doesn't matter whether we use h or $\hbar/2$ or, for that matter, $100h$. All that matters is that we use a constant with the dimensions of position $\times$ momentum, and h is a convenient choice. We'll explain why at the end of Section 5.2.

For one particle confined in a box of volume V, the "small" number of cells in phase space around energy E is then

$$\mathrm{d}\Omega_{\text{tot}} \approx \frac{[x][p_x]}{h}\,\frac{[y][p_y]}{h}\,\frac{[z][p_z]}{h} = \frac{V}{h^3}\,[p_x][p_y][p_z], \tag{3.5}$$

because $[x], [y], [z]$ range over the whole dimensions of the box, so their product is V. The $[p_x], [p_y], [p_z]$ are very small ranges around the nominal values of p_x, p_y, p_z. The total number of cells in phase space for this particle for all energies $0 \to E$ is Ω_{tot}:

$$\Omega_{\text{tot}}(E) \approx \frac{V}{h^3} \times \text{a volume of momentum space in 3 dimensions.} \tag{3.6}$$

For N particles with *total* energy E, the total number of states up to energy E is

$$\Omega_{\text{tot}}(E) = \frac{V^N}{h^{3N}} \times \text{a volume of momentum space in } 3N \text{ dimensions.} \tag{3.7}$$

Label the particles' momenta $p_{1x}, p_{1y}, \ldots, p_{Nz}$. Then $p_{1x}^2 + \cdots + p_{Nz}^2 = 2mE$ where m is the particle mass. So we require the volume of a sphere in $3N$ dimensions with radius $\sqrt{2mE}$. The volume of a sphere of radius R in n dimensions is

$$\text{volume} = \frac{\pi^{n/2} R^n}{(n/2)!}. \tag{3.8}$$

(Try this formula for $n = 1, 2, 3$, using $(1/2)! = \sqrt{\pi}/2$.)

On a side note, the factorial is actually defined as a function over the complex numbers. An alternative notation for it is $\Pi(z) \equiv z!$, which lends itself to writing its derivative $\Pi'(z)$. You will usually see the equivalent notation $\Gamma(z+1) \equiv z!$ in textbooks. Why is there a "$+1$" there? There is no good reason for this, and I think we would all be better off by dropping it once and for all. The gamma function with its unnecessary $+1$ was even declared outdated in favour of the factorial by Sir M.J. Lighthill a half century ago in his classic text on Fourier analysis [2]. If you do use the gamma function you will find yourself writing tedious expressions like $\Gamma(a+1) = \Gamma(b+1)\,\Gamma(c+1)$, and you will forever need to remind yourself of the $+1$ when calculating simple things like $\Gamma(5)$. Strike a blow for notational simplicity and write $\Pi(z)$ when you need to use function notation for the factorial.

Setting $n = 3N$ and $R = \sqrt{2mE}$ in (3.8), then writing $\gamma \equiv 3N/2$, (3.7) becomes

$$\Omega_{\text{tot}}(E) = \frac{V^N}{h^{2\gamma}} \frac{\pi^\gamma (2mE)^\gamma}{\gamma!} \,. \tag{3.9}$$

This is the total number of states up to energy E. We get at $\Omega(E)$, the number of states *at* energy E, via the density of states $g(E)$ using (3.4):

$$g(E) = \Omega'_{\text{tot}}(E) = \frac{V^N}{h^{2\gamma}} \frac{(2\pi m)^\gamma}{\gamma!} \gamma E^{\gamma-1} \,. \tag{3.10}$$

With N large, use Stirling's rule:

$$g(E) \approx \frac{V^N}{h^{2\gamma}} \frac{(2\pi m)^\gamma e^\gamma \gamma}{\gamma^{\gamma+1/2}\sqrt{2\pi}} E^{\gamma-1} \approx V^N \left(\frac{2\pi m\,e\,E}{h^2 \gamma}\right)^\gamma. \tag{3.11}$$

The main result here is that $g(E) \propto V^N E^{3N/2-1} \approx V^N E^{3N/2}$ when N is large. If we do require the number of states "at" a particular energy E, we can estimate it using (3.3) with a suitable choice of ΔE. But it turns out that we won't have to do this.

Do note that a consequence of the above continuum approximation is that the number of states $\Omega(E)$ and their density $g(E)$ are often treated somewhat interchangeably in statistical mechanics. We'll explain why this is done at the end of Section 5.2.

Example 1. We have a cubic room of side 5 m with $N = 10^{27}$ distinguishable particles at 300 K, each with mass equal to the average mass of an air molecule (4.8×10^{-26} kg). What is the density of states $g(E)$? Use the Equipartition Theorem (proved later) to set a value for E of $N\,3kT/2$.

> Notational device: write $a \stackrel{b}{=} \equiv a \times 10^b$.

$$\log_{10} g(E) = 1.5 \stackrel{27}{=} \log_{10} \left(\frac{25 \times 2\pi \times 4.8 \stackrel{-26}{=\!=} \times 2.7 \times {}^3\!/_2 \times 1.38 \stackrel{-23}{=\!=} \times 300 \times 10^{27}}{\left(6.63 \stackrel{-34}{=\!=}\right)^2 \times 1.5 \stackrel{27}{=\!=}} \right)$$
$$\simeq 3.5 \stackrel{28}{=\!=}\,. \tag{3.12}$$

So

$$g(E) \approx 10^{3.5\times10^{28}} \text{ states/joule.} \qquad \underline{\text{Answer}} \tag{3.13}$$

How big is this number? If we just settle for writing it out as approximately 1 followed by a string of 0s, with each 0 being 1 cm across, then the length of this string will be about 37 thousand million light years, or several times the extent of the observable universe. That's not *how big* the number is; rather, that's just how big its decimal representation is. The number itself is *stupendously* bigger.

Identical particles

To do the same analysis for identical particles, realise that if the number of states is much larger than the number of particles (which is very true above), the above calculation will over-count by a factor of $N!$. So divide the result in (3.13) by $10^{27}!$ (that's a factorial, and is left as an exercise) to obtain

$$g(E) \approx 10^{8.4 \times 10^{27}} \text{ states/joule.} \qquad \underline{\text{Answer}} \qquad (3.14)$$

If the number of states is not $\gg$ the number of particles, we will need the approach used in quantum statistics later in the course.

3.2 Density of States for More Complicated Structures

Consider the variables that quantify a substance's energy. There is a background potential energy u_0 along with various potential and kinetic energies, so the energy of *one* particle is ε for the following regimes:

Monatomic liquid and gas:

$$\varepsilon = u_0 + \frac{p_x^2}{2m} + \cdots + \frac{p_z^2}{2m}. \qquad (3.15)$$

In liquids, u_0 is complicated as molecular configurations fluctuate rapidly. In gases, $u_0 \simeq 0$.

Solid crystal lattice: particles are like harmonic oscillators, with x, y, z measuring their displacement from equilibrium:

$$\varepsilon = u_0 + \frac{kx^2}{2} + \cdots + \frac{kz^2}{2} + \frac{p_x^2}{2m} + \cdots + \frac{p_z^2}{2m}. \qquad (3.16)$$

Gas of complex molecules: with angular momenta $L_1, \ldots, L_3$ about principal axes, paired with moments of inertia $I_1, \ldots, I_3$, along with reduced mass μ, vibration frequency ω, vibrational separation r:

$$\varepsilon = u_0 + \underbrace{\frac{p_x^2}{2m} + \ldots}_{\text{translation}} + \underbrace{\frac{L_1^2}{2I_1} + \ldots}_{\text{rotation}} + \underbrace{\frac{1}{2}\mu \dot{r}^2 + \frac{1}{2}\mu\omega^2 r^2}_{\text{vibration}}. \qquad (3.17)$$

Each of these variables that contributes to the energy via a square is called a *degree of freedom*. (Why are the two terms for vibration above called two degrees of freedom when one cannot be changed without also changing the other? The name is something of a misnomer; a degree of freedom is simply defined as a term that contributes to the energy via a square. Both of those terms do that.)

If we re-derive $g(E)$ for these more complex structures, the same general arguments apply, but with some modifications as follows. Each extra degree of freedom contributes an extra dimension per particle to the sphere in phase space. E.g., if the number of degrees of freedom per particle is $\nu = 5$ (for a diatomic molecule, since experiments show that this

won't spin around its main axis), the "$3N/2$" above is replaced by $5N/2$ and the various coefficients change. Also, correcting for identical particles (when there are far more states than particles, which is the case here) means dividing by $N!$, or approximately by N^N. In that case, for an ideal gas we can write expressions for the number of accessible states for distinguishable and for identical particles:

$$\Omega_{\text{dist}} \propto V^N \left(\frac{E}{N}\right)^{\nu N/2}, \quad \Omega_{\text{ident}} \propto \left(\frac{V}{N}\right)^N \left(\frac{E}{N}\right)^{\nu N/2}. \tag{3.18}$$

The particles of a *solid* are not free to move about their container like a gas, so there is no spatial volume term. Also, only the *thermal energy* $\varepsilon - u_0$ contributes to the number of accessible states. So for a solid we write

$$\Omega_{\text{dist}} \propto \left(\frac{E - Nu_0}{N}\right)^{\nu N/2}. \tag{3.19}$$

The particles of a solid are certainly distinguishable by their locations at the various lattice sites, so there is no Ω_{ident} to be considered.

4 Zeroth and First Laws of Thermodynamics

When two systems interact, energy can be transferred in three ways:

(a) transfer of heat: "thermal" (conduction, convection, radiation),

(b) doing work: "mechanical" (pressure, fields),

(c) letting particles move: "diffusive" (permeable membranes).

Zeroth Law of Thermodynamics

> If two systems are in thermal/mechanical/diffusive equilibrium with a third system, then they're in thermal/mechanical/diffusive equilibrium with each other.

4.1 Preparing for the First Law of Thermodynamics

The First Law of Thermodynamics is a statement of the conservation of energy. It is usually—perhaps always—expressed in terms of infinitesimals, so we will first make some comments about these.

Infinitesimal quantities, also called differentials, are used extensively in statistical mechanics. What does a quantity like dt mean? Consider deriving a particle's velocity $v(t)$ from its position $s(t)$ in one dimension. We might Taylor-expand $s(t + \Delta t)$ to write

$$v(t) \equiv \lim_{\Delta t \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{s(t) + s'(t)\,\Delta t + \frac{1}{2!}s''(t)\,\Delta t^2 + \cdots - s(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} s'(t) + \frac{1}{2!} s''(t) \, \Delta t + \dots$$

$$= s'(t) \,. \tag{4.1}$$

But we could just as well write all of this as

$$v(t) = \frac{s(t + \mathrm{d}t) - s(t)}{\mathrm{d}t} = \frac{s(t) + s'(t) \, \mathrm{d}t - s(t)}{\mathrm{d}t} = s'(t) \,. \tag{4.2}$$

This last expression is an economical and elegant way of writing the previous one. By writing "$\mathrm{d}t$", we really mean "$\Delta t + O(\Delta t^2)$" along with a statement of an eventual division by Δt and a limit being taken as $\Delta t \to 0$. So when speaking of an infinitesimal, or an "infinitesimally small quantity", we are really referring to the end result of a limit process applied to the non-infinitesimal Δt.

> This sort of idea also applies to the *delta function* used widely in Fourier analysis. The delta function $\delta(x)$ is usually defined as an infinitely tall spike at $x = 0$ and zero elsewhere, with $\int_{-\infty}^{\infty} \delta(x) \, \mathrm{d}x \equiv 1$. Any expression involving this function can be treated as the limit of a sequence of similar expressions that each replace the delta by a bell-shaped function, where these bell-shaped functions become increasingly narrower and higher in the limit.
>
> Treating delta functions and infinitesimals as tied to a limit process gives them a firm foundation, although with delta functions we need always to ask whether it's valid to swap the relevant manipulation and the limit process. You will find the occasional book stating that infinitesimals need advanced ideas of differential geometry to give them substance. If you do investigate further to analyse what this might mean, I suggest that you'll only find notation that became briefly fashionable some decades ago but never went anywhere, presumably because what it was designed to do was able to be done more simply in other ways. The kernel of what infinitesimals are all about is contained in (4.1) and (4.2). Infinitesimals as defined by equations like these are certainly used routinely in differential geometry, where they have a natural and very central role. So, rather than say infinitesimals need differential geometry to give them meaning, I would choose to say that infinitesimals help to give differential geometry meaning.

Differentials are Increases

Correctly translating a physics task into the language of mathematics goes a long way to making it tractable. In particular for doing problems in thermodynamics, we'll stress the following point. For *any* quantity f, the symbols Δf and $\mathrm{d}f$ refer to *increases* in f, meaning $f_{\text{final}} - f_{\text{initial}}$. Similarly, $-\Delta f$ and $-\mathrm{d}f$ are *decreases* in f. Usually Δf and $\mathrm{d}f$ are called "changes in f", but this is not a very useful phrase if we give "change" its everyday meaning of the absolute value of increase or decrease. After all, discarding a sign is not a good idea! Likewise, when we write ∇f and ∂f in the context of partial derivatives, we are referring indirectly to increases in f as other quantities are increased. And remember that an increase can be negative—that's what is meant by a decrease. The same idea holds for vectors too: by $\Delta \boldsymbol{v}$ we mean $\boldsymbol{v}_{\text{final}} - \boldsymbol{v}_{\text{initial}}$. This is the increase in $\boldsymbol{v}$, although the idea of a vector increasing might not be as intuitive as it is for everyday numbers, because the *length* of $\boldsymbol{v}$ needn't change when $\boldsymbol{v}$ increases. But that's okay; after all, who said anything about length? Just remember that

$$\Delta \;=\; \text{increase} \;=\; \text{final} - \text{initial} \,, \tag{4.3}$$

which applies to vectors as well as numbers, and you won't go wrong. (It *is* still meaningful to refer generically to a "change in a quantity" if we're not concerned with its value.)

If we keep the correct language in mind, we will have no problems recognising that a term like $-\mathrm{d}V$ means a decrease in volume: when a volume V gets smaller, the amount by which it decreased, $-\mathrm{d}V$, is positive. For example, when we push a piston to squeeze the air in a cylinder, we do work equal to the pressure P that we applied (always positive) times the loss in volume $-\mathrm{d}V$ (again positive), or $-P\mathrm{d}V$; thus the energy E of the gas increases by this amount: $\mathrm{d}E = -P\mathrm{d}V$. The pressure changes as the volume decreases, which is why we write this using infinitesimals. We could also write

$$\Delta E = \int_{V_i}^{V_f} -P\mathrm{d}V, \tag{4.4}$$

which means neither more nor less than the infinitesimal expression (but takes more room to write!). Keeping clear in our minds whether the quantities appearing in the equations of statistical mechanics are increasing or decreasing helps us relate the mathematics to the physics. A good example of such care will appear in Section 9.3.

Exact and Inexact Differentials

A differential $\mathrm{d}f$ is called *exact* if the state of a system possesses a unique value of f.

As an example, suppose two people ("1" and "2") walk from Adelaide to Melbourne. They follow different paths and meet at some point en route. This point has a particular height h above sea level. The position of each walker always has a unique value of h, and so $\mathrm{d}h$ is an exact differential. If we write a walker's height $h(\alpha, \omega)$ as a function of latitude α and longitude ω, then $\Delta h = \int \mathrm{d}h$ is independent of the path that each took to get to their meeting point. Also, Δh equals "final height minus initial height".

Their meeting point has a particular latitude α and longitude ω, and so $\mathrm{d}\alpha, \mathrm{d}\omega$ are also exact differentials. An exact differential like $\mathrm{d}h$ can be written as $\mathrm{d}h = A\,\mathrm{d}\alpha + B\,\mathrm{d}\omega$, which is entirely equivalent to writing

$$\frac{\partial h}{\partial \alpha} = A, \quad \frac{\partial h}{\partial \omega} = B. \tag{4.5}$$

Contrast the walkers' height h at their meeting point by the distances they each have walked. Their common position doesn't possess a unique value of a variable called distance s. Certainly s is well defined for each walker, but in general they have walked different distances s_1, s_2.

Now suppose each walker takes an infinitesimal step, increasing s_1 and s_2 by infinitesimal amounts. These steps are certainly well defined, but are not exact differentials because the state called "position" does not have a unique s. The steps are written $\text{đ}s_1, \text{đ}s_2$, and are called *inexact differentials*. We can use a generic step $\text{đ}s$ to write the total distances covered by each walker on arriving in Melbourne as

$$s_1 = \int_{\text{path 1}} \text{đ}s, \quad \text{and } s_2 = \int_{\text{path 2}} \text{đ}s. \tag{4.6}$$

Notice that we might not be inclined to write their total distances covered as $\Delta s_1, \Delta s_2$. That's because Δs means "final s minus initial s", but there *is* no variable s that is a function of position. Perhaps we could put a bar through the Δ symbol, but this is not something that anyone does.

Another example of an inexact differential is the small area element used to discuss Gauss's theorem in electromagnetism. This is usually written dA, but it's not as if we have an area A that's increasing by dA. This area element is more properly written đA. Likewise, the force due to air pressure that acts on a small surface of area đA could be written đF but it's more usually written dF. Perhaps the use of đ is confined to statistical mechanics, but there's no real reason why this should be so.

The Inexact Differential "Heat in", đQ

Suppose we are given a container of hot gas. The subject of thermodynamics deals with the processes this gas might've undergone to bring it to its present state. It might've been heated over a stove ("thermal" in the Zeroth Law), by doing work on it ("mechanical"), by changing its chemical environment ("diffusive"), or some combination of the three. Knowing nothing of the gas's history, we cannot generally ascertain how it was heated. Heating it over a stove gives it thermal energy Q, but doing work on it gives it mechanical energy $\int -P \, dV$, which manifests as the same heat. Both of these actions lead to the same final state, so we cannot say the gas has a unique "heat" Q associated with it. We *can*, however, talk about a small amount of heat being put into the gas, and this must then be an inexact differential, written đQ. And just as in the discussion immediately following (4.6) of the distance s covered by a walker, we will always write Q, never ΔQ, for a large amount of heat put into a system.

Although we have written the mechanical work done on the gas as $-P \, dV$, it's sometimes introduced by writing it as đW. Again the bar is necessary because the state does not have a unique parameter called work W associated with it; any work we do on the system must be an inexact differential. But notice that this inexact differential đW can actually be written as an exact differential $-P \, dV$, because V is certainly a state variable! We'll show later that the same can be done for đQ. This rewriting of đQ in terms of a new state variable was a key discovery of statistical mechanics, and was the central idea that allowed thermodynamics to be analysed and extended using statistical mechanics ideas.

First Law of Thermodynamics (Provisional Form)

The infinitesimal increase in internal energy of a system is given by thermal, mechanical, and diffusive contributions as follows:

$$dE = đQ - P \, dV + \mu \, dN. \tag{4.7}$$

đQ = heat put **in** by e.g. a stove.

$-P \, dV$ = pressure $\times$ loss in volume = work done **on** system.

$\mu \, dN$ = *chemical potential* $\times$ increase in particle number = energy brought **in** by new particles that isn't related to heat transfer or work. Due to new environment created by incoming particles. (Think of adding water to a concentrated acid: it heats up dangerously.)

The term $-P \, dV$ is just one example of work being done on the system. Others exist,

such as an electric field which can change dipole moments, giving a term $-\boldsymbol{E}\cdot\mathrm{d}\boldsymbol{p}$. We will write $-P\,\mathrm{d}V$ to represent all such terms.

Our primary goal will be to rewrite the đQ in the First Law as an exact differential, since the relevant quantities can then be used to quantify the state of a system. We'll take V as one of these quantities from the outset (so won't stop to write đW), and will see later how to replace đQ by something else.

4.2 Partial Derivatives and Variables Held Constant

A good understanding of partial derivatives is useful in statistical mechanics. Here are some points worth noting.

Consider a function $f(x, y, z)$. When writing

$$\frac{\partial f(x, y, z)}{\partial x} \quad \text{or} \quad \left(\frac{\partial f}{\partial x}\right)_{y,\,z} \tag{4.8}$$

we mean $\mathrm{d}f/\mathrm{d}x$ at constant y and z. Thus

$$\left(\frac{\partial x}{\partial f}\right)_{y,\,z} = \frac{1}{(\partial f/\partial x)_{y,\,z}} \,. \tag{4.9}$$

However, normally when we swap the roles of e.g. f and x, the set of variables that are being held fixed actually changes, and so the simple reciprocation of (4.9) doesn't apply. A more familiar example relates polar coordinates to cartesians. Begin with

$$x = r\cos\theta\,, \quad y = r\sin\theta\,. \tag{4.10}$$

In such an arena when we write something like $\partial x/\partial r$, we mean $(\partial x/\partial r)_\theta$; that is, differentiate with respect to one variable (r) holding all others of its family (θ) constant. With this convention, the set of partial derivatives of one set of coordinates with respect to the other is usually written as the elements of a matrix, called a *jacobian matrix*:

$$\begin{bmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\[2mm] \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}. \tag{4.11}$$

There are two jacobian matrices: one has the partial derivatives of cartesians with respect to polars, and the other has the partial derivatives of polars with respect to cartesians. Now watch carefully: by multiplying these two matrices, you should be able to see that the following relationship holds:

$$\begin{bmatrix} \dfrac{\partial r}{\partial x} & \dfrac{\partial r}{\partial y} \\[2mm] \dfrac{\partial \theta}{\partial x} & \dfrac{\partial \theta}{\partial y} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\[2mm] \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\[2mm] -\dfrac{\sin\theta}{r} & \dfrac{\cos\theta}{r} \end{bmatrix}. \tag{4.12}$$

This is an important relationship because it enables us to invert partial derivatives when the set of variables being held constant switches from one set of coordinates to the other. For example, comparing the "1,1" elements of (4.11) and (4.12) shows that

$$\left(\frac{\partial x}{\partial r}\right)_\theta = \left(\frac{\partial r}{\partial x}\right)_y = \cos\theta\,. \tag{4.13}$$

With the above convention of omitting the constant variables in mind, this is normally just written

$$\frac{\partial x}{\partial r} = \frac{\partial r}{\partial x} = \cos\theta\,. \tag{4.14}$$

This might at first look a little odd, until we realise that each derivative holds a different variable constant, and so simple reciprocation cannot be used.

Example 1: Show that $(\partial x/\partial r)_\theta = \frac{1}{(\partial r/\partial x)_\theta}$.

Start with $r = x/\cos\theta$, so that

$$\left(\frac{\partial r}{\partial x}\right)_\theta = \frac{1}{\cos\theta} = \frac{1}{(\partial x/\partial r)_\theta}\,. \qquad \underline{\underline{\text{QED}}} \tag{4.15}$$

Example 2: What is $(\partial\theta/\partial r)_y$?

The easiest approach is to differentiate both sides of $y = r\sin\theta$ with respect to r, holding y constant, to get $0 = \sin\theta + r\cos\theta\,(\partial\theta/\partial r)_y$. Or, for a slight variation, calculate $(\partial r/\partial\theta)_y$ in the same way and form the reciprocal. Here is a third way, which might give you more insight. Draw two infinitesimally separated points at constant y. One has polar coordinates (r,θ); the other has $(r + \mathrm{d}r, \theta + \mathrm{d}\theta)$. Noting that we need keep only the lowest powers necessary of the infinitesimals, write

$$\begin{aligned} y = r\sin\theta &= (r + \mathrm{d}r)\,\sin(\theta + \mathrm{d}\theta)\\ &= (r + \mathrm{d}r)\,(\sin\theta + \cos\theta\,\mathrm{d}\theta)\\ &= r\sin\theta + \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta\,. \end{aligned} \tag{4.16}$$

So at constant y we have $-\sin\theta\,\mathrm{d}r = r\cos\theta\,\mathrm{d}\theta$, in which case

$$\left(\frac{\partial\theta}{\partial r}\right)_y = \frac{-\sin\theta}{r\cos\theta}\,. \qquad \underline{\underline{\text{Answer}}} \tag{4.17}$$

Note that although we seemed to work to first order only, the answer is *exact*. Can you see why? If not, study (4.1) and (4.2) in the more familiar language of position $s(t)$ and velocity $v(t)$ in one dimension, and remember that velocity is a first-order increase in position with respect to time.

5   Accessible States for Interacting Systems

We wish to focus on the elusive notion of heat transfer. So consider 2 systems interacting thermally, but not mechanically or diffusively. They're isolated, so their total energy $= E$, a constant. System 1 has N_1 particles, each with ν_1 d.o.f. and total energy E_1. Similarly for system 2. What is the total number of accessible states Ω as a function of E_1?

$$
\begin{array}{cc}
N_1,\, \nu_1 & N_2,\, \nu_2 \\
\underline{\;\; E_1 \;\;} & \underline{\;\; E_2 = E - E_1 \;\;} \\
\text{system 1} & \text{system 2}
\end{array}
$$

Define $\gamma_1 \equiv \nu_1 N_1/2$ and $\gamma_2 \equiv \nu_2 N_2/2$, so that the results of Section 3 give

$$\Omega_1 \propto E_1^{\gamma_1}, \quad \Omega_2 \propto E_2^{\gamma_2}. \tag{5.1}$$

Then

$$\Omega(E_1) = \Omega_1\,\Omega_2 \propto E_1^{\gamma_1}\,(E - E_1)^{\gamma_2}, \quad 0 \leqslant E_1 \leqslant E. \tag{5.2}$$

Consider plotting $\Omega(E_1)$-vs-E_1. The stationary points occur when $\Omega'(E_1) = 0$:

$$\Omega'(E_1) = E_1^{\gamma_1 - 1}\,(E - E_1)^{\gamma_2 - 1}\left[\gamma_1 E - (\gamma_1 + \gamma_2)E_1\right] \overset{\text{req.}}{=\!=\!=} 0. \tag{5.3}$$

So stationary points occur at $E_1 = 0$ and E (minima), and $E_1 = \widehat{E}_1 \equiv \gamma_1 E/(\gamma_1 + \gamma_2)$ (maximum).

How wide is the peak at $E = \widehat{E}_1$? A useful measure is α such that

$$\Omega(\widehat{E}_1 + \alpha\widehat{E}_1) = \frac{1}{2}\,\Omega(\widehat{E}_1), \tag{5.4}$$

so that $2\alpha\widehat{E}_1$ is approximately the "full width at half maximum" (commonly known as the FWHM). (This is only approximate, as the peak isn't necessarily symmetric.) In (5.2) we see that Ω is a product and involves powers, so its logarithm turns out to be simpler to work with; also, a logarithm will lead to a better approximation of the peak when we Taylor-expand shortly, because the logarithm of a strongly peaked function is not strongly peaked and so needs fewer Taylor terms to describe it. So we introduce a new symbol which will play a central role in statistical mechanics, called the *statistical entropy*:

$$\boxed{\sigma \equiv \ln \Omega.} \tag{5.5}$$

Use this to write (5.2) as

$$
\begin{aligned}
\sigma(E_1) &= \text{constant} + \gamma_1 \ln E_1 + \gamma_2 \ln(E - E_1), \\
\text{so } \sigma'(E_1) &= \frac{\gamma_1}{E_1} - \frac{\gamma_2}{E - E_1}, \\
\text{and } \sigma''(E_1) &= \frac{-\gamma_1}{E_1^2} - \frac{\gamma_2}{(E - E_1)^2}.
\end{aligned}
\tag{5.6}
$$

[We can't "really" take the logarithm of a dimensioned number such as energy; but the constant of proportionality in (5.2) effectively introduces a scaling factor for the units that does allow us to take a log. But this constant has no effect on the physical arguments

and is perhaps a little tedious to include everywhere. See the comment just after (9.2).]
Equation (5.4) becomes

$$\sigma(\widehat{E}_1 + \alpha\widehat{E}_1) = -\ln 2 + \sigma(\widehat{E}_1)\,, \tag{5.7}$$

which Taylor-expands to

$$\sigma'(\widehat{E}_1)^{\nearrow 0}\,\alpha\widehat{E}_1 + \sigma''(\widehat{E}_1)\,\frac{\alpha^2\widehat{E}_1^{\,2}}{2} \simeq -\ln 2\,. \tag{5.8}$$

A little work gives

$$\alpha \simeq \sqrt{\frac{2\,\gamma_2\ln 2}{\gamma_1(\gamma_1 + \gamma_2)}}\,. \tag{5.9}$$

When $\gamma_1 = \gamma_2 \simeq 10^{24}$, we get $\alpha \simeq 10^{-12}$, and so the FWHM is about $2 \times 10^{-12}\,\widehat{E}_1$. This is tiny compared with $\widehat{E}_1$, so the thermal interaction means that systems 1 and 2 are extremely likely to have energies

$$\widehat{E}_1 = \frac{\gamma_1 E}{\gamma_1 + \gamma_2} = \frac{\nu_1 N_1\,E}{\nu_1 N_1 + \nu_2 N_2}\,,$$

$$\widehat{E}_2 \equiv E - \widehat{E}_1 = \frac{\gamma_2 E}{\gamma_1 + \gamma_2} = \frac{\nu_2 N_2\,E}{\nu_1 N_1 + \nu_2 N_2} \tag{5.10}$$

respectively.

Fluctuations By what factor f does the number of accessible states Ω drop if E_1 should exceed $\widehat{E}_1$ by one part per million? That is, calculate

$$f \equiv \frac{\Omega(\widehat{E}_1)}{\Omega\big((1 + 10^{-6})\widehat{E}_1\big)}\,. \tag{5.11}$$

Taylor-expand $\ln f$:

$$\ln f = \sigma(\widehat{E}_1) - \sigma\big(\widehat{E}_1 + 10^{-6}\widehat{E}_1\big) \simeq -\sigma''(\widehat{E}_1)\,\frac{10^{-12}\,\widehat{E}_1^{\,2}}{2} \simeq 10^{12}\,. \tag{5.12}$$

Hence

$$f \simeq e^{10^{12}} \simeq 10^{0.4343 \times 10^{12}} = 10^{434,300,000,000}\,. \qquad \underline{\underline{\text{Answer}}} \tag{5.13}$$

This is a huge drop. So, with the system equally likely to be in any of its accessible states (by postulate), then the chance of a 1 ppm fluctuation away from energies $\widehat{E}_1, \widehat{E}_2$ is so minute that we can discount it from *ever* happening. (We should really do an integral here to consider a fluctuation of *at least* 1 ppm, but the above calculation serves to give a good idea of the numbers involved.)

5.1 Defining Temperature, and the Equipartition Theorem

Prior to the advent of statistical mechanics, the concept of temperature was already known in a heuristic way from thermodynamics. An early success of statistical mechanics was its

precise definition of temperature using the above idea of two thermally interacting systems. At *thermal equilibrium* when the energy has distributed itself as $\widehat{E}_1, \widehat{E}_2$, we *define* the two systems to have equal values of temperature T. This is done by our noticing in (5.10) that the average energy per particle per degree of freedom is the same for both systems, so that the common value can be used to define their temperature:

$$\frac{\widehat{E}_1}{\nu_1 N_1} = \frac{\widehat{E}_2}{\nu_2 N_2} = \frac{E}{\nu_1 N_1 + \nu_2 N_2} \equiv \frac{kT}{2} \,. \tag{5.14}$$

Here k is *Boltzmann's constant*, inserted to allow this statistical definition of temperature to be equated with the everyday thermodynamic idea of temperature (as we'll soon see; so we'll assume from now on that temperature is a positive quantity). The factor of 2 ensures compatibility with other uses of Boltzmann's constant. In practice we must provide something extra to disentangle temperature from k. This is done by setting $T = 273.16\,\mathrm{K}$ at the triple point of water ($\simeq 0.01^\circ\mathrm{C}$).

> Note that the SI unit of the Kelvin temperature scale is a "kelvin", not a "Kelvin degree". A temperature of $100\,\mathrm{K}$ is vocalised "one hundred kelvins"—not "one hundred degrees Kelvin", nor "one hundred Kelvin". In common with all SI units, "kelvin" is written with a lowercase k but, being someone's name, its short form uses an uppercase K.

The two systems can be intermixed. E.g., system 1 might refer to translation ($\nu_1 = 3$) of all the (diatomic) molecules present, and system 2 might refer to their rotation ($\nu_2 = 2$). So (5.14) says that the translational degrees of freedom possess energy $3kT/2$ per particle, and the rotational degrees of freedom possess energy $2kT/2$ per molecule. This constitutes the *Equipartition Theorem*.

Equipartition Theorem

> If the equilibrium distribution is
>
> – the most probable distribution consistent with constant total energy and constant particle number, and
>
> – there is no restriction on the number of particles in any one state, and
>
> – thermal energy E varies continuously with a coordinate u, and
>
> – E depends on u^2,
>
> then the energy associated with this coordinate is $kT/2$.

5.2 Entropy and the Second Law of Thermodynamics

Equation (3.10) implies that when N is large, $\Omega = f(V, N)\, E^{\nu N/2}$ for some function f. That means

$$\sigma(E, V, N) = \ln \Omega = \ln f(V, N) + \frac{\nu N}{2} \ln E \,. \tag{5.15}$$

Thus

$$\left(\frac{\partial \sigma}{\partial E} \right)_{V,N} = \frac{\nu N}{2E} \xlongequal{\text{Equipartition}} \frac{\nu N}{2\,\nu N kT/2} = \frac{1}{kT} \,. \tag{5.16}$$

Define the *thermodynamic entropy* (usually just called entropy) of a general system to be $S \equiv k\sigma$, i.e.

$$S = k\ln\Omega\,. \tag{5.17}$$

With this definition, (5.16) becomes

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N}\,, \quad \text{or} \quad T = \left(\frac{\partial E}{\partial S}\right)_{V,N}\,. \tag{5.18}$$

This is often taken as the definition of T.

Entropy is *additive*, since for any two systems 1 and 2 with entropies $S_1 = k\ln\Omega_1$ and $S_2 = k\ln\Omega_2$, the entropy of the combined system *before they interact* is $k\ln(\Omega_1\,\Omega_2) = k\ln\Omega_1 + k\ln\Omega_2 = S_1 + S_2$.

Since a system composed of two subsystems having attained equilibrium is overwhelmingly likely to be found in a state for which Ω, or S, is maximised, we can say that

> When two systems interact, the entropy of the combined system increases along the path to equilibrium.

This is one statement of the Second Law of Thermodynamics. There are several others that are equivalent on various levels.

The Use of Planck's Constant for Quantifying Entropy

Near the start of Section 3.1 we said that it doesn't matter whether the volume of one cell of phase space is defined by dividing by h for each dimension, or $\hbar/2$ or $100h$. The reason is because throughout our study of entropy in statistical mechanics, it will only ever be an *increase* in entropy, ΔS, that has a role in the calculations. Even when we eventually write S by itself in (8.2), it is still only ΔS that ever matters.

With that in mind, consider how entropy relates to a volume $\mathcal{V}$ of phase space:

$$\Delta S = S_f - S_i = k\ln\Omega_f - k\ln\Omega_i = k\ln\frac{\Omega_f}{\Omega_i} = k\ln\frac{\mathcal{V}_f/h^{\nu N}}{\mathcal{V}_i/h^{\nu N}} = k\ln\frac{\mathcal{V}_f}{\mathcal{V}_i}\,. \tag{5.19}$$

It's apparent that we could replace h by any multiple of h and nothing would change in the last equation: ΔS would still only be determined by a ratio of phase space volumes. In fact, h isn't needed at all. As used in Section 3.1 it was really only a device giving us a way of specifying and counting states for a continuous system, which is a modern way of approaching entropy. An alternative approach might define entropy through the idea of phase space volume alone, but that would divorce entropy from the idea of the number of accessible states. Defining entropy via the number of accessible states allows us to build an intuition about it, because we can then consider very simple discrete systems and count their states easily.

This reasoning also explains why, as we noted on page 9, the number of states $\Omega(E)$ and their density $g(E)$ are often treated as interchangeable in statistical mechanics. Although entropy is defined as the logarithm of $\Omega(E)$, we usually only have knowledge of $g(E)$. However, (3.3) ties these together by way of some unspecified energy width ΔE, which

then acts as a factor in the mathematics without changing the physics, just like any factor we might choose to put in front of Planck's constant. So this ΔE is not usually written explicitly. A slight complication is that unlike a factor in front of Planck's constant, ΔE has a dimension; but any choice of units will do because they're all related by scaling factors, which ultimately vanish because we only ever really consider ΔS. In fact, any absence of ΔE might be considered offset by the fact that we wrote λ for $\lambda - 1$ in (3.11). However, (3.11) used Stirling's rule, which itself is just an approximation! You can see that there is some vagueness in the number of factors of energy when we are dealing with large numbers of particles.

5.3   Heat, Entropy, and the First Law Again

When examining the energy of a system as a function of some variables, we have axes V (volume, corresponding to mechanical interactions) and N (particle number, corresponding to diffusive interactions). But a third axis is needed to account for thermal interactions: the $đQ$ term in the First Law. That's why we considered a thermal-only interaction at the start of Section 5. Heating the system corresponds to increasing the entropy S, which means we can choose to make the third axis simply entropy since, unlike heat, entropy is a state variable like V and N. In that case

$$dE = \frac{\partial E}{\partial S}\,dS - P\,dV + \mu\,dN. \tag{5.20}$$

For a quasistatic process we have $T = (\partial E/\partial S)_{V,N}$, producing

$$\boxed{dE = T\,dS - P\,dV + \mu\,dN.} \tag{5.21}$$

Comparing this with the provisional form of the First Law (4.7) allows us to write $đQ = T\,dS$, irrespective of whether V, N are constant or not. Note that for the interacting systems 1 and 2 at the start of Section 5 on their way to equilibrium, no heat went in from the outside. But their combined entropy went up! So on the way to equilibrium, $đQ = 0$ but $dS > 0$. We can't write $đQ = T\,dS$ for the *combined* system; what would T be here, since it's only defined at equilibrium? But certainly at equilibrium T *is* defined, and then $đQ = T\,dS = 0$. In general, $đQ = T\,dS$ is only written for quasistatic processes, since these are always arbitrarily close to equilibrium. Since we only consider such processes, (5.21) can be considered as the final form of the First Law of Thermodynamics, and we will use it extensively.

Directions of Flow from the First Law

Consider systems 1 and 2 interacting thermally, mechanically, and diffusively:

$$dS = dS_1 + dS_2. \tag{5.22}$$

Suppose energy, volume, and particle number are conserved. Express dS_2 in terms of dS_1 using the First Law:

$$dE_2 = -dE_1\,, \quad \text{so} \quad T_2\,dS_2 - P_2\,dV_2 + \mu_2\,dN_2 = -T_1\,dS_1 + P_1\,dV_1 - \mu_1\,dN_1\,,$$

$$\left.\begin{array}{l} \mathrm{d}V_2 = -\mathrm{d}V_1 \\ \mathrm{d}N_2 = -\mathrm{d}N_1 \end{array}\right\} \text{ so } \quad T_2\,\mathrm{d}S_2 + P_2\,\mathrm{d}V_1 - \mu_2\,\mathrm{d}N_1 = -T_1\,\mathrm{d}S_1 + P_1\,\mathrm{d}V_1 - \mu_1\,\mathrm{d}N_1\,. \quad (5.23)$$

Thus

$$\mathrm{d}S_2 = \frac{-T_1\,\mathrm{d}S_1}{T_2} + \frac{(P_1 - P_2)\,\mathrm{d}V_1}{T_2} + \frac{(\mu_2 - \mu_1)\,\mathrm{d}N_1}{T_2}\,, \quad (5.24)$$

so that (5.22) becomes

$$\mathrm{d}S = \left(1 - \frac{T_1}{T_2}\right)\mathrm{d}S_1 + \frac{P_1 - P_2}{T_2}\,\mathrm{d}V_1 + \frac{\mu_2 - \mu_1}{T_2}\,\mathrm{d}N_1\,. \quad (5.25)$$

As the system heads toward equilibrium, $\mathrm{d}S > 0$. We have the freedom to control how much of each interaction in (5.25) occurs. So we require the entropy to increase for each interaction if that interaction occurs on its own. Hence consider each term on the right hand side of (5.25) separately to conclude:

Thermal:

$$\left(1 - \frac{T_1}{T_2}\right)\mathrm{d}S_1 > 0\,, \text{ so } \begin{cases} T_1 < T_2 \text{ and } \mathrm{d}S_1 > 0 \ (\text{\dj}Q_1 > 0) \\ \qquad\qquad \text{or} \\ T_1 > T_2 \text{ and } \mathrm{d}S_1 < 0 \ (\text{\dj}Q_1 < 0) \end{cases} \quad (5.26)$$

So heat flows toward the region of lower temperature.

Mechanical:

$$\frac{P_1 - P_2}{T_2}\,\mathrm{d}V_1 > 0\,, \text{ so } \begin{cases} P_1 > P_2 \text{ and } \mathrm{d}V_1 > 0 \\ \qquad\qquad \text{or} \\ P_1 < P_2 \text{ and } \mathrm{d}V_1 < 0 \end{cases} \quad (5.27)$$

So the boundary moves toward the region of lower pressure.

Diffusive:

$$\frac{\mu_2 - \mu_1}{T_2}\,\mathrm{d}N_1 > 0\,, \text{ so } \begin{cases} \mu_2 > \mu_1 \text{ and } \mathrm{d}N_1 > 0 \\ \qquad\qquad \text{or} \\ \mu_2 < \mu_1 \text{ and } \mathrm{d}N_1 < 0 \end{cases} \quad (5.28)$$

So particles flow toward the region of lower chemical potential.

At equilibrium, as before, each term $= 0$ individually. Hence the temperatures, pressures, and chemical potentials become equal.

A Note on Heat Flow   Equation (5.24) ($\times\, T_2$) says

$$\text{\dj}Q_2 = -\text{\dj}Q_1 + \text{terms involving } \mathrm{d}V_1, \mathrm{d}N_1\,. \quad (5.29)$$

That is, the heat flowing into system 2 will only equal the heat flowing out of 1 if the interaction is purely thermal.

Intensive and Extensive Variables

T, P, μ are *intensive*: they are only defined at equilibrium, and they don't scale with the system; they are constant throughout the system.

S, V, N are *extensive*: they are defined even outside of equilibrium, and they scale *linearly* with the system.

In the First Law these variables occur in *conjugate pairs*, with each term being an intensive variable times an infinitesimal increase in its conjugate, extensive, partner. Experimentally, these two types of variable seem to be sufficient to quantify all systems in statistical mechanics.

5.4 "Deriving" the Ideal Gas Law

Consider rearranging the First Law to give

$$\mathrm{d}S = \frac{\mathrm{d}E}{T} + \frac{P\,\mathrm{d}V}{T} - \frac{\mu\,\mathrm{d}N}{T}\,. \tag{5.30}$$

Thus, in particular,

$$\frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{E,N}. \tag{5.31}$$

But we know from (3.18) that for an ideal gas, $\Omega = f(E, N)\,V^N$ for some function f. Its entropy is thus $S = k \ln f(E, N) + Nk \ln V$. Substitute this into (5.31) to give

$$\frac{P}{T} = \frac{Nk}{V}\,, \tag{5.32}$$

which is the ideal gas law:

$$\boxed{PV = NkT\,.} \tag{5.33}$$

Actually, we haven't really *derived* the ideal gas law here. What we have really done is shown that our statistical definition of temperature is consistent with the thermodynamic definition of temperature.

With the number of moles $n \equiv N/N_A$ (where N_A = Avogadro's number) and the *gas constant* $R \equiv N_A k \simeq 8.314$ SI units, we obtain the molar form of the ideal gas law: $PV = nRT$. It's often convenient in chemical calculations to replace k by R/N_A: the gas constant R is a conveniently simple number, and N_A allows well-known molar quantities to be used. Good examples of this are found in Section 9.

Example: What is the volume of $1\,\mathrm{kg}$ of O_2 gas at 1 atmosphere and $20^\circ\mathrm{C}$ ($101{,}325\,\mathrm{Pa}$ and $293.15\,\mathrm{K}$)?

Since one mole of O_2 has a mass of about $32\,\mathrm{g}$, we are dealing with $n = 1000/32$ moles of what is essentially an ideal gas. Then the required volume is

$$V = \frac{nRT}{P} \simeq \frac{1000 \times 8.314 \times 293.15}{32 \times 101{,}325}\ \mathrm{m}^3 = 0.752\ \mathrm{m}^3 = 752\,\ell\,. \ \underline{\underline{\text{Answer}}} \tag{5.34}$$

6   Heat Capacity

Define the *heat capacity* of a substance, when heated while holding parameter A constant, as the heat added divided by the temperature increase it produces: $C_A \equiv đQ/dT$. Normally particle number N is constant, so

$$C_A \, dT = đQ = dE + P \, dV. \tag{6.1}$$

At constant volume, $C_V \, dT = dE$, or

$$C_V = \left(\frac{\partial E}{\partial T}\right)_{V,N}. \tag{6.2}$$

At constant pressure, for an ideal gas we have

$$\begin{aligned}
C_P \, dT &= dE + P \, dV \\
&= C_V \, dT + P \, dV \\
&= C_V \, dT + nR \, dT.
\end{aligned} \tag{6.3}$$

That means

$$C_P = C_V + nR = C_V + Nk. \tag{6.4}$$

Now define

$$\text{molar heat capacity } C^{\text{mol}} \equiv C/n\,,$$
$$\text{specific heat capacity } C^{\text{sp}} \equiv C/m \quad (m = \text{total mass of the substance}). \tag{6.5}$$

Equation (6.4) leads to a useful expression in chemistry:

$$C_P^{\text{mol}} = C_V^{\text{mol}} + R. \tag{6.6}$$

Given that specific heat capacity is usually fairly constant over everyday temperature ranges of interest, the expression $đQ = C_A \, dT = m C_A^{\text{sp}} \, dT$ can be integrated to give the total heat energy Q that must be absorbed by a mass m to increase its temperature by ΔT:

$$\boxed{Q \simeq m C_A^{\text{sp}} \, \Delta T\,.} \tag{6.7}$$

The specific heat capacity is usually just called specific heat.

6.1   The Adiabatic Process

This is a process for which $đQ = 0$: no heat is exchanged between the system and its environment. Particle number N is usually taken as constant, so write $dE + P \, dV = 0$. Then, for an ideal gas,

$$C_V \, dT + P \, dV = 0\,,$$
$$\text{or } C_V \, \frac{d(PV)}{nR} + P \, dV = 0\,.$$
$$\text{That means } C_V \, dP\, V + C_V \, P \, dV + nR \, P \, dV = 0\,,$$

$$\text{or } C_V V \, \mathrm{d}P + (C_V + nR) \, P \mathrm{d}V = 0 \,,$$
$$\text{so } C_V V \, \mathrm{d}P + C_P \, P \mathrm{d}V = 0 \,. \tag{6.8}$$

Divide by PVC_V:

$$\frac{\mathrm{d}P}{P} = \frac{-C_P}{C_V} \frac{\mathrm{d}V}{V} \,. \tag{6.9}$$

Define a (temperature-dependent) parameter γ:

$$\gamma \equiv \frac{C_P}{C_V} = \frac{C_P^{\mathrm{mol}}}{C_V^{\mathrm{mol}}} = \frac{C_P^{\mathrm{sp}}}{C_V^{\mathrm{sp}}} \,. \tag{6.10}$$

In that case

$$\ln P = -\gamma \ln V + \text{const.}$$
$$\text{or } P \propto V^{-\gamma} \,,$$
$$\text{so } \boxed{PV^{\gamma} = \text{constant.}}$$

This last expression is used when describing adiabatic processes for ideal gases on a PV diagram. The adiabatic process, together with the isothermal process ($PV = $ constant), make up the Carnot cycle that forms the core description of the thermodynamics of engines and refrigerators.

For an ideal gas, $E = \nu N k T / 2$, so $C_V = \nu N k / 2$. Then

$$\gamma = \frac{C_V + Nk}{C_V} = 1 + \frac{Nk}{\nu Nk/2} = \frac{\nu + 2}{\nu} \,. \tag{6.11}$$

So measuring C_P and C_V gives information on the structure of the gas molecules:

$$\nu = \frac{2}{\gamma - 1} \,. \tag{6.12}$$

Example: The entropy of water at $25°$C and 1 atmosphere is $188.8 \, \mathrm{J\,K^{-1}mol^{-1}}$. Water's molecular mass is 0.018 kg/mol and its specific heat is $4186 \, \mathrm{J\,K^{-1}kg^{-1}}$. Raise its temperature to $27°$C. What is the new molar entropy? (This example comes from page 183 #40 of [1].)

Deal with one mole. We require S, its entropy at $27°$C. When heating anything, it's wise not to bolt the lid down, so we assume the above specific heat is for constant pressure, although what is being held constant actually doesn't matter for the purpose of the calculation (i.e., the number 4186 has the process built into it, and we don't need to know how the temperature was increased). We can only assume the specific heat is constant over temperature. The entropy increase from $25°$C is

$$\Delta S = \int \mathrm{d}S = \int \frac{\mathrm{d}Q}{T} = \int_{25°\mathrm{C}}^{27°\mathrm{C}} \frac{C \, \mathrm{d}T}{T} = C \ln \frac{273.15 + 27}{273.15 + 25} \,, \tag{6.13}$$

where C is the heat capacity of 1 mole; this is the heat capacity of 0.018 kg, so that $C = 4186 \, \mathrm{J\,K^{-1}kg^{-1}} \times 0.018$ kg. Then

$$S = 188.8 \text{ J/K} + \Delta S$$

$$= 188.8 + 4186 \times 0.018 \, \ln \frac{300.15}{298.15} \quad \text{J/K}$$

$$= 189.3 \, \text{J/K} \,. \tag{6.14}$$

So the new molar entropy is $189.3 \, \text{J\,K}^{-1}\text{mol}^{-1}$.   <u>Answer</u>

Third Law of Thermodynamics

Experimentally, there seems never to be more than one state available ultimately as a system's temperature $\to 0$. This suggests the Third Law of Thermodynamics:

> In the limit of temperature going to zero, a system's entropy goes to zero, regardless of its makeup or the makeup of its environment.

This allows us to write a system's entropy at some temperature T and constant parameter(s) A explicitly as

$$S_A(T) = S_A(T) - S_A(0) = \int_{T=0}^{T} \mathrm{d}S_A = \int_0^T \frac{C_A \, \mathrm{d}T}{T} \,. \tag{6.15}$$

7   The Flow of Heat Energy

The current density $\boldsymbol{J}$ (also known as *flux density*) of heat energy is defined as the vector pointing in the direction of heat flow, whose length is the energy crossing a perpendicular unit area in unit time in that direction. Experimentally it's found to be proportional to the spatial rate of loss of T, or $-\nabla T$. So

$$\boldsymbol{J} = -\kappa \, \nabla T \,, \tag{7.1}$$

where $\kappa > 0$ is the *thermal conductivity*. The *heat current* across an area A is

$$I = \int \boldsymbol{J} \cdot \boldsymbol{n} \, \mathrm{d}A = -\kappa \int \nabla T \cdot \boldsymbol{n} \, \mathrm{d}A \,, \tag{7.2}$$

where the unit vector $\boldsymbol{n}$ is perpendicular to $\mathrm{d}A$. But for any scalar function T, the increase in T along a small step $\boldsymbol{n} \, \mathrm{d}\ell$ in space is $\mathrm{d}T = \nabla T \cdot \boldsymbol{n} \, \mathrm{d}\ell$. Then

$$\frac{\mathrm{d}T}{\mathrm{d}\ell} \text{ in the } \boldsymbol{n} \text{ direction} = \nabla T \cdot \boldsymbol{n} \tag{7.3}$$

(called a *directional derivative*, and sometimes written $\partial T/\partial n$.) Equation (7.2) becomes

$$I = -\kappa \int \left(\frac{\mathrm{d}T}{\mathrm{d}\ell} \text{ along } \boldsymbol{n} \right) \mathrm{d}A$$

$$= -\kappa \left\langle \frac{\mathrm{d}T}{\mathrm{d}\ell} \, \substack{\perp \text{ to} \\ \text{surface}} \right\rangle A \,, \tag{7.4}$$

where $\langle \cdot \rangle$ denotes the mean value over the area A. So

$$\left\langle \frac{-\mathrm{d}T}{\mathrm{d}\ell} \, \substack{\perp \text{ to} \\ \text{surface}} \right\rangle = \frac{I}{\kappa A} \,, \tag{7.5}$$

which leads to

$$\left\langle -\Delta T \underset{\text{surface}}{\perp \text{ to}} \right\rangle \simeq I \underbrace{\frac{\Delta \ell}{\kappa A}}_{\equiv R} \, . \qquad (7.6)$$

Here R is the *thermal resistance*, analogous to electrical resistance by way of Ohm's Rule, which says that an electric current I experiences a drop in electric potential Φ across a resistance R of

$$-\Delta \Phi = IR \, . \qquad (7.7)$$

(The drop $-\Delta \Phi$ is more usually written V.) When connecting thermal resistances in series and parallel to model heat flow through complex objects, we add them in the same way as we do electrical resistances. The quantity $RA = \Delta \ell / \kappa$ is called the *R-factor* in the building trade.

Example: An $18\,\text{m} \times 6\,\text{m}$ roof is made of $25\,\text{mm}$ thick pine board with thermal conductivity $\kappa = 0.11\,\text{W m}^{-1}\,\text{K}^{-1}$, covered with asphalt shingles of R-factor $R_f = 0.0776\,\text{K m}^2\,\text{W}^{-1}$. Neglecting the overlap of the shingles, how much heat is conducted through the roof when the inside temperature is 21°C and the outside temperature is 5°C?

When speaking of conducted heat, we mean the heat current I, measured in watts. We must calculate, using (7.6),

$$I = \frac{-\Delta T}{R_{\text{pine}} + R_{\text{asph}}} = \frac{-\Delta T}{\frac{\Delta \ell}{\kappa A}(\text{pine}) + \frac{R_f}{A}(\text{asph})}$$

$$= \frac{-\Delta T \, A}{\Delta \ell / \kappa (\text{pine}) + R_f (\text{asph})} = \frac{16 \times 18 \times 6}{0.025/0.11 + 0.0776} = 5.67\,\text{kW}. \qquad \underline{\text{Answer}}$$

7.1 The Continuity Equation

This is a general equation describing *local conservation* of some quantity. Consider an energy density ϱ_E in a volume V. There is a current density $\boldsymbol{J}$ carrying energy across the surface A of the volume. In a time $\text{d}t$, the volume loses an amount of energy equal to $-\text{d} \int \varrho_E \, \text{d}V$. This equals the amount that flows out through the closed surface, which is $\oint \text{d}t \, \boldsymbol{J} \cdot \boldsymbol{n} \, \text{d}A$. So

$$-\text{d} \int \varrho_E \, \text{d}V = \oint \text{d}t \, \boldsymbol{J} \cdot \boldsymbol{n} \, \text{d}A = \text{d}t \int \nabla \cdot \boldsymbol{J} \, \text{d}V \, . \qquad (7.8)$$

That means

$$\int \text{d}V \left[\frac{\partial \varrho_E}{\partial t} + \nabla \cdot \boldsymbol{J} \right] = 0 \quad \text{for all volumes } V. \qquad (7.9)$$

Since the volume is quite arbitrary, we arrive at the *continuity equation*:

$$\boxed{\frac{\partial \varrho_E}{\partial t} + \nabla \cdot \boldsymbol{J} = 0 \, .} \qquad (7.10)$$

This idea of local conservation contrasts with *global conservation*, in which a quantity might vanish at one point but re-appear at another. Although the quantity might well have been conserved, there may have been no flow across any surface in between the two points of vanishing and emergence. This is a weak form of the idea of conservation; local conservation is a much stronger concept. For example, energy is always found to be conserved locally.

7.2   The Heat Equation (a.k.a. Diffusion Equation)

Since $\boldsymbol{J} = -\kappa\,\nabla T$, the continuity equation becomes

$$\frac{\partial \varrho_E}{\partial t} - \kappa\,\nabla^2 T = 0\,. \tag{7.11}$$

But an increase in energy $\mathrm{d}E$ in the volume equals $mC^{\mathrm{sp}}\,\mathrm{d}T$ where the mass inside is m. So divide $\mathrm{d}E = mC^{\mathrm{sp}}\,\mathrm{d}T$ by the volume to get $\mathrm{d}\varrho_E = \varrho_m C^{\mathrm{sp}}\,\mathrm{d}T$ where ϱ_m is the mass density. Thus

$$\frac{\partial \varrho_E}{\partial t} = \varrho_m C^{\mathrm{sp}}\,\frac{\partial T}{\partial t}\,. \tag{7.12}$$

Put this into (7.11) to get

$$\boxed{\nabla^2 T = \frac{\varrho_m C^{\mathrm{sp}}}{\kappa}\,\frac{\partial T}{\partial t}\,.} \tag{7.13}$$

This is the *heat equation*, or *diffusion equation*. It has been produced by combining the continuity equation with the experimental observation that the current density is proportional to the spatial loss in T.

7.3   Solving the Heat Equation

The general heat equation is

$$\nabla^2 T = \frac{1}{K}\,\frac{\partial T}{\partial t} \quad (K > 0)\,. \tag{7.14}$$

We can see why this might well model the flow of heat. The reason is because $\nabla^2 T$ is a second spatial derivative. When there are no sources and T is peaked (a hot spot), the second spatial derivative is negative, which means $\partial T/\partial t$ is also negative. So the temperature in a hot spot decreases with time. Similarly, when T is a trough (a cold spot), the second spatial derivative is positive, which means $\partial T/\partial t$ is also positive. That means the temperature in a cold spot increases with time. This behaviour is just what we expect of temperature.

The heat equation is linear, from which it follows that any linear combination of solutions is also a solution. There is a huge literature devoted to solving partial differential equations that relate the laplacian operator ∇^2 to zeroth, first, and second time derivatives. The topic is normally found in applied maths courses, so here we'll consider just one approach useful for dealing with an infinite domain (i.e., no boundary conditions). One solution to the heat equation is

$$T(t, \boldsymbol{x}) = \frac{1}{(4\pi K t)^{3/2}}\,\exp\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{4Kt}\,. \quad \text{(Prove it!)} \tag{7.15}$$

This is a (normalised) gaussian of width $\sigma = \sqrt{2Kt}$. When $t = 0$ this becomes a delta function, $\delta(\boldsymbol{x} - \boldsymbol{x}')$. That is,

$$T(0, \boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}') \Longrightarrow T(t, \boldsymbol{x}) = \frac{1}{(4\pi K t)^{3/2}}\,\exp\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{4Kt}\,. \tag{7.16}$$

Now, suppose the initial temperature distribution is known to be $T(0, \boldsymbol{x}) = f(\boldsymbol{x})$. We can certainly write

$$f(\boldsymbol{x}) = \int_{-\infty}^{\infty} f(\boldsymbol{x}')\, \delta(\boldsymbol{x} - \boldsymbol{x}')\, \mathrm{d}^3 x'\,. \tag{7.17}$$

We have expressed the initial temperature distribution as a linear combination of delta functions, and each of these delta functions evolves according to (7.16). That means the general solution of the heat equation is

$$T(t, \boldsymbol{x}) = \int_{-\infty}^{\infty} f(\boldsymbol{x}')\, \underbrace{\frac{1}{(4\pi K t)^{3/2}}\, \exp \frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{4K t}}_{\textit{Green function for heat equation}}\, \mathrm{d}^3 x'\,. \tag{7.18}$$

So provided we can do the integral, perhaps numerically, any initial temperature distribution can be propagated forwards in time.

Equation (7.18) has the form

$$f(x) = \int_{-\infty}^{\infty} a(x')\, b(x - x')\, \mathrm{d}x' \equiv a(x) * b(x)\,, \tag{7.19}$$

which is called the *convolution* of $a(x)$ and $b(x)$. In this language, (7.18) is written

$$T(t, \boldsymbol{x}) = T(0, \boldsymbol{x}) * \frac{1}{(4\pi K t)^{3/2}}\, \exp \frac{-|\boldsymbol{x}|^2}{4K t}\,. \tag{7.20}$$

At each moment in time, the convolution is essentially a moving spatial mean: it acts to smear out the initial temperature distribution, with the (gaussian) smear getting wider and wider as time goes on, just as we expect will happen as heat flows. The theory of convolutions is intimately related to Fourier and Laplace transforms.

8 Integrating the Internal Energy

Consider

$$E = \int \mathrm{d}E = \int (T\,\mathrm{d}S - P\,\mathrm{d}V + \mu\,\mathrm{d}N)\,. \tag{8.1}$$

Because T, P, μ are intensive and S, V, N are extensive, we can calculate E by breaking the *final* complete system up into many small pieces, each with T, P, μ at their final values, and each piece having $\mathrm{d}S, \mathrm{d}V, \mathrm{d}N$. Then simply add these to get

$$E = TS - PV + \mu N\,. \tag{8.2}$$

It might seem surprising that we have been able to integrate the energy so easily, but the method works precisely because of the distinction between intensive and extensive variables in the First Law.

Calculating $\mathrm{d}E$ once more from (8.2) and then applying the First Law produces the *Gibbs–Duhem equation*

$$S\,\mathrm{d}T - V\,\mathrm{d}P + N\,\mathrm{d}\mu = 0\,, \tag{8.3}$$

which relates changes in the intensive variables. It also shows that if one of these intensive variables changes, then at least one other intensive variable must change.

8.1   Switching Dependence on Variables

How to switch from a dependence on e.g. S to its conjugate variable T? Define the *Helmholtz Free Energy* $F \equiv E - TS$. (It is called free energy because in a heat engine it is the energy available, or "free", to do work.)

$$\mathrm{d}F = \mathrm{d}E - S\,\mathrm{d}T - T\,\mathrm{d}S = -S\,\mathrm{d}T - P\mathrm{d}V + \mu\,\mathrm{d}N\,. \tag{8.4}$$

Compare this to the First Law: we have swapped the S and T by way of defining a new variable F. (This technique of defining a new variable by adding or subtracting the product of the relevant conjugate pair is an example of a *Legendre transform*, which appears in other areas of physics.) For an **isothermal nondiffusive** process, $\mathrm{d}T = \mathrm{d}N = 0$, so

$$\mathrm{d}F = -P\mathrm{d}V = \text{work done on system}. \tag{8.5}$$

Similarly, to interchange P and V, define the *enthalpy* $H \equiv E + PV$.

$$\mathrm{d}H = T\,\mathrm{d}S + V\,\mathrm{d}P + \mu\,\mathrm{d}N\,. \tag{8.6}$$

Isobaric nondiffusive processes are of great relevance to chemical reactions, which are often performed in an open vessel and so are isobaric. For these,

$$\mathrm{d}H = T\,\mathrm{d}S = \text{heat entering system}. \tag{8.7}$$

Enthalpy tells us whether such reactions make their surroundings hotter or colder. The reaction $a \to b$ has a total heat energy entering the system of $\Delta H = H_b - H_a$. When $\Delta H < 0$ the reaction is *exothermic*: the reaction vessel gets hotter. When $\Delta H > 0$ the reaction is *endothermic*: the reaction vessel gets colder.

Define the *Gibbs Free Energy* $G \equiv E - TS + PV = \mu N$.

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}P + \mu\,\mathrm{d}N\,. \tag{8.8}$$

For an **isothermal isobaric** process, $\mathrm{d}G = \mu\,\mathrm{d}N$ (the $N\,\mathrm{d}\mu$ is absent on account of the Gibbs–Duhem equation). So the Gibbs free energy is useful in analysing diffusive processes.

The *Grand Free Energy* $\Phi \equiv E - TS - \mu N = -PV$ can also be defined, with

$$\mathrm{d}\Phi = -S\,\mathrm{d}T - P\mathrm{d}V - N\,\mathrm{d}\mu\,. \tag{8.9}$$

This is useful for describing systems with constant T, V, μ (isothermal, isochoric, nondiffusive).

8.2   Maxwell Relations

These are simply expressions of mixed partial derivatives from the First Law. Consider a function $f(x, y, z)$. We have

$$\mathrm{d}f = X\,\mathrm{d}x + Y\,\mathrm{d}y + Z\,\mathrm{d}z\,. \tag{8.10}$$

But we know that e.g.

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y}\,; \quad \text{i.e. } \frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x}\,, \tag{8.11}$$

or, more explicitly,

$$\left(\frac{\partial X}{\partial y}\right)_{x,z} = \left(\frac{\partial Y}{\partial x}\right)_{y,z}. \tag{8.12}$$

Apply this idea to the First Law: $\mathrm{d}E = T\,\mathrm{d}S - P\,\mathrm{d}V + \mu\,\mathrm{d}N$. Then e.g.

$$\left(\frac{\partial T}{\partial V}\right)_{S,N} = -\left(\frac{\partial P}{\partial S}\right)_{V,N}, \tag{8.13}$$

and so on. This is an example of a Maxwell relation. Others follow from F, G, H. For example, using $\mathrm{d}F = -S\,\mathrm{d}T - P\,\mathrm{d}V + \mu\,\mathrm{d}N$, we get

$$-\left(\frac{\partial P}{\partial N}\right)_{T,V} = \left(\frac{\partial \mu}{\partial V}\right)_{T,N} \tag{8.14}$$

and so on. These relations are useful in allowing us to change variables depending on the experimental setup. It's usually best to use as independent variables ones that are either constrained or easily measured. Because three variables are needed to describe a system, we can choose one of each type (thermal, mechanical, diffusive). E.g., what is the dependence of E on T, P for a closed system? (We can ignore μ, N here as there is no diffusion.) We have

$$\mathrm{d}E = \left(\frac{\partial E}{\partial T}\right)_P \mathrm{d}T + \left(\frac{\partial E}{\partial P}\right)_T \mathrm{d}P. \tag{8.15}$$

We can consider E as a function of S, V, each a function of T, P, to write

$$\left(\frac{\partial E}{\partial T}\right)_P = \left(\frac{\partial E}{\partial S}\right)_V \left(\frac{\partial S}{\partial T}\right)_P + \left(\frac{\partial E}{\partial V}\right)_S \left(\frac{\partial V}{\partial T}\right)_P. \tag{8.16}$$

But $\mathrm{d}E = T\,\mathrm{d}S - P\,\mathrm{d}V$, so $(\partial E/\partial S)_V = T$ and $(\partial E/\partial V)_S = -P$. Thus (8.16) becomes

$$\left(\frac{\partial E}{\partial T}\right)_P = T\left(\frac{\partial S}{\partial T}\right)_P - P\left(\frac{\partial V}{\partial T}\right)_P. \tag{8.17}$$

Similarly, $(\partial E/\partial P)_T$ can be found.

An example of dependencies that are easily measured starts with

$$T = \left(\frac{\partial E}{\partial S}\right)_{V,N}. \tag{8.18}$$

This leads to (assuming here that N is constant, so we'll drop reference to N):

$$\begin{aligned}
C_V &= \left(\frac{\mathrm{d}Q}{\mathrm{d}T}\right)_V = \left(\frac{T\,\mathrm{d}S}{\mathrm{d}T}\right)_V = T\left(\frac{\partial S}{\partial T}\right)_V, \\
C_P &= \left(\frac{\mathrm{d}Q}{\mathrm{d}T}\right)_P = \left(\frac{T\,\mathrm{d}S}{\mathrm{d}T}\right)_P = T\left(\frac{\partial S}{\partial T}\right)_P.
\end{aligned} \tag{8.19}$$

Also define

$$\text{coeff. of thermal expansion } \beta \equiv \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P$$
$$= \text{relative increase in } V \text{ with } T \text{ at constant } P, \tag{8.20}$$

and

$$\text{coeff. of isothermal compressibility } \kappa \equiv \frac{-1}{V} \left(\frac{\partial V}{\partial P} \right)_T$$

$$= \text{relative decrease in } V \text{ with } P \text{ at constant } T. \tag{8.21}$$

Another useful tool arises from answering a question such as: if $z = z(x, y)$ and z is held constant, how do x and y relate? Write

$$0 = \mathrm{d}z = \left(\frac{\partial z}{\partial x} \right)_y \mathrm{d}x + \left(\frac{\partial z}{\partial y} \right)_x \mathrm{d}y. \tag{8.22}$$

Hence

$$-\left(\frac{\partial z}{\partial x} \right)_y \mathrm{d}x = \left(\frac{\partial z}{\partial y} \right)_x \mathrm{d}y, \tag{8.23}$$

in which case

$$\left(\frac{\partial x}{\partial y} \right)_z = \frac{(\partial z/\partial y)_x}{-(\partial z/\partial x)_y} = -\left(\frac{\partial z}{\partial y} \right)_x \left(\frac{\partial x}{\partial z} \right)_y. \tag{8.24}$$

For example, at constant N,

$$\frac{\beta}{\kappa} = -\left(\frac{\partial V}{\partial T} \right)_P \left(\frac{\partial P}{\partial V} \right)_T = \left(\frac{\partial P}{\partial T} \right)_V. \tag{8.25}$$

Although a study of the Maxwell Relations forms a useful exercise in dealing with the various partial derivatives that arise in statistical mechanics, they do not form a major part of the modern subject, and we will not pursue them further.

9 The Chemical Potential and Phase Changes

Consider interactions of various substances. For there to be much interaction at all, the particles should be able to move about; so we will model the number of accessible states of a complex material to have a volume term just like a gas of identical particles, combining the ideas of (3.18) and (3.19):

$$\Omega \propto \left(\frac{V}{N} \right)^N \left(\frac{E - N u_0}{N} \right)^{\nu N/2}. \tag{9.1}$$

Set $\varrho \equiv N/V =$ particle density. Also, Equipartition says $E - N u_0 = \nu N k T/2$. So

$$\Omega \propto \varrho^{-N} T^{\nu N/2}. \tag{9.2}$$

The implied normalisation here will allow Ω to be dimensionless, and can be embodied in a sort of "reference density" ϱ^* and "reference temperature" T^*. In that case the entropy $S = k \ln \Omega$ would be written

$$S = \text{const.} - Nk \ln(\varrho/\varrho^*) + \nicefrac{1}{2} \nu Nk \ln(T/T^*), \tag{9.3}$$

since we can only take the logarithm of a dimensionless quantity. But in practice, ϱ^* and T^* will end up cancelling out anyway, so we will simply absorb them into ϱ and T, and write the entropy as

$$S = \text{const.} - Nk \ln \varrho + \nicefrac{1}{2} \nu Nk \ln T. \tag{9.4}$$

Use this to calculate μ:

$$\mu N = E + PV - TS = \underbrace{Nu_0 + {}^1\!/\!_2\, \nu NkT}_{E} \; + \; \underbrace{\left\{ \begin{array}{c} NkT \text{ (gas)} \\ \text{negligible (liq./sol.)} \end{array} \right\}}_{PV}$$

$$\underbrace{- \text{const.} \times T + NkT \ln \varrho - {}^1\!/\!_2\, \nu NkT \ln T}_{-TS}, \qquad (9.5)$$

or

$$\mu = u_0 + kT \ln \varrho + g(T) \quad \text{for some function } g. \qquad (9.6)$$

That is, seeking a lower μ goes with seeking regions of lower "ambient" energies u_0 and seeking regions of lower particle densities.

9.1 Colligative Properties of Solutions

These are properties that only depend on the concentration of a dissolved substance (the *solute*), but not on what it actually is. We will use the example of adding a small amount of common salt (the solute) to pure water (the *solvent*), which approximates sea water. Sea water typically has 2% salt by particle number, meaning there are approximately 2 salt ions (we're not concerned with whether they are Na^+ or Cl^-) to every 98 water ions (which could be H_3O^+ or OH^-). For this case, set a variable $f = 0.02$.

Initially (before adding salt), the water has density $\varrho_0 = N/V$.

Finally (salt + water), the water's density is $\varrho = (1 - f)N/V = (1 - f)\varrho_0$.

Focus on the water's chemical potential: for small f it increases by

$$\Delta\mu = \mu_f - \mu_i \simeq u_0 + kT \ln \varrho + g(T)$$
$$- u_0 - kT \ln \varrho_0 - g(T)$$
$$= kT \ln \frac{\varrho}{\varrho_0} = kT \ln(1 - f) \simeq -fkT. \qquad (9.7)$$

[This is a good example of how ϱ^* cancels: we might choose to make the replacements $\varrho \to \varrho/\varrho^*$, $\varrho_0 \to \varrho_0/\varrho^*$, but that won't change the last line of (9.7).]

Start with (1) water in equilibrium with its vapor (both with chemical potential μ), then (2) salt is added, which reduces the chemical potential of the water to $\mu - fkT$ and so destroys the equilibrium, then (3) liquid/vapor equilibrium is again restored, with a new chemical potential μ'. There have been overall increases in temperature and pressure of $\Delta T, \Delta P$. Now make use of the Gibbs–Duhem equation (8.3), writing that equation as

$$\Delta\mu \simeq \frac{-S}{N}\Delta T + \frac{V}{N}\Delta P. \qquad (9.8)$$

Also set the entropy per particle to be $s \equiv S/N$, and the volume per particle to be $v \equiv V/N$. Now we note the following:

For the vapor, in going from (1) to (3) in the previous paragraph,

$$\Delta\mu = \mu' - \mu \simeq -s_{\text{vap}}\Delta T + v_{\text{vap}}\Delta P. \qquad (9.9)$$

For the water, in going from (1) to (2), there are approximately no changes in temperature and pressure (this is a simplification), and in going from (2) to (3),

$$\Delta\mu = \mu' - (\mu - fkT) \simeq -s_{\text{liq}}\Delta T + v_{\text{liq}}\Delta P\,. \tag{9.10}$$

Combine (9.9), (9.10) to give

$$-s_{\text{vap}}\Delta T + v_{\text{vap}}\Delta P + fkT \simeq -s_{\text{liq}}\Delta T + v_{\text{liq}}\Delta P\,, \tag{9.11}$$

or

$$-\left(s_{\text{vap}} - s_{\text{liq}}\right)\Delta T + \left(v_{\text{vap}} - v_{\text{liq}}\right)\Delta P \simeq -fkT\,. \tag{9.12}$$

Examine this last equation for the two cases of constant temperature and constant pressure:

Constant Temperature

$$\left(v_{\text{vap}} - v_{\text{liq}}\right)\Delta P \simeq -fkT\,. \tag{9.13}$$

But $v_{\text{vap}} \gg v_{\text{liq}}$, so write, for the vapor, $v\Delta P \simeq -fkT$, and treating the vapor as an ideal gas (so $Pv = kT$), this becomes

$$\frac{-\Delta P}{P} \simeq f\,. \tag{9.14}$$

That is, the relative drop in vapor pressure $\simeq f$, the fraction of solute particles in solution.

Constant Pressure

$$\left(s_{\text{vap}} - s_{\text{liq}}\right)\Delta T \simeq fkT\,. \tag{9.15}$$

Follow n particles leaving the liquid: they have picked up the latent heat of vaporisation, which increases the distance between particles without increasing their kinetic energy (and therefore without increasing their temperature). They each have mass m, and the specific latent heat of vaporisation (i.e. per unit mass) is L^{sp}. So

$$\begin{aligned}
s_{\text{vap}} - s_{\text{liq}} &= s \text{ for the } n \text{ particles after entering the vapor} \\
&\quad - s \text{ for the } n \text{ particles before leaving the liquid} \\
&= \frac{\Delta S}{n} = \frac{Q}{Tn}, \text{ (where } Q = \text{latent heat in)} \\
&= \frac{L^{\text{sp}}nm}{Tn}\,. \tag{9.16}
\end{aligned}$$

Thus (9.15) becomes

$$\frac{L^{\text{sp}}m}{T}\Delta T \simeq fkT\,, \quad \text{so} \quad \Delta T \simeq \frac{fkT^2}{L^{\text{sp}}m}\,. \tag{9.17}$$

If the latent heat is specified as a molar latent heat L^{mol} (latent heat per mole), then the last equation becomes

$$\Delta T \simeq \frac{fRT^2}{L^{\text{sp}}mN_A} = \frac{fRT^2}{\text{lat. heat for 1 mol}} = \frac{fRT^2}{L^{\text{mol}}}\,. \tag{9.18}$$

Raising Boiling Points We add salt to water. That lowers the chemical potential of the water by fkT, so that the vapor has a higher chemical potential. Thus vapor particles start entering the brine (because diffusion always occurs from high chemical potential to low chemical potential). Consider the water initially boiling, in equilibrium with its vapor. Add salt, and vapor starts to enter the brine. If we wish to restore the equilibrium at constant pressure, we must increase the temperature by ΔT. That is, the boiling point of the brine will now be $100^\circ\text{C} + \Delta T$.

Example: Calculate the temperature at which the above 2% salt water mixture boils. Its molar latent heat of vaporisation is $L^{\text{mol}} = 40{,}700$ J/mol.

Use $f = 0.02$ and write

$$\Delta T = \frac{fRT^2}{L^{\text{mol}}} = \frac{0.02 \times 8.314 \times 373^2}{40{,}700} \simeq 0.6\,\text{K}. \tag{9.19}$$

So the new boiling point is 100.6°C.   <u>Answer</u>

Lowering Melting Points Replace the water vapor in the above discussions with ice. Now, instead of vapor particles entering the brine, ice particles enter the brine. That is, when salt is added to a water/ice mixture, the ice starts to melt. To restore equilibrium at constant pressure, we must now *remove* heat. This is equivalent to making L^{mol} negative in the above equations. If the molar latent heat of fusion of water is 6000 J/mol, the new melting point of the ice is $0^\circ\text{C} + \Delta T$, where

$$\Delta T = \frac{fRT^2}{L^{\text{mol}}} = \frac{0.02 \times 8.314 \times 273^2}{-6000} \simeq -2\,\text{K}. \tag{9.20}$$

So the new freezing point of the brine is -2°C. Because adding salt to an ice/water mixture causes the ice to melt, this is one method of de-icing roads in winter—provided the ambient temperature is not too low.

9.2 Osmotic Pressure

Suppose some water (μ) is separated from brine ($\mu - fkT$) by a membrane through which water can pass, but not salt. Imagine this was originally 2 systems, both at μ. Now we lower the brine's potential to $\mu - fkT$. Gibbs–Duhem says that this change goes together with a ΔP and a ΔT:

$$\Delta\mu = -fkT = -s_{\text{brine}}\Delta T + v_{\text{brine}}\Delta P. \tag{9.21}$$

At constant temperature, there will now be an *osmotic pressure* "forcing" water molecules to diffuse from μ to $\mu - fkT$ (high to low potentials):

$$\Delta P = \left(\frac{N}{V}\right)_{\text{brine}} \Delta\mu. \tag{9.22}$$

We know the pressure acts to force water into the brine, so consider absolute values, writing (with V_{mol} the volume of one mole, and there are N particles of water in the brine)

$$\text{osmotic pressure} = \left(\frac{N}{V}\right)_{\text{brine}} fkT = \frac{N\,fkT}{\frac{N}{N_A}V_{\text{mol}}} = \frac{fRT}{V_{\text{mol}}}. \tag{9.23}$$

Example: What is the osmotic pressure for pure water diffusing into brine at $25°$C?

$$\text{pressure} = \frac{0.02 \times 8.314 \times 298}{18 \underline{\underline{-6}}} \text{ Pa} = 2.75 \text{ MPa} \simeq 27 \text{ atm.} \qquad \underline{\underline{\text{Answer}}} \qquad (9.24)$$

This is very large! The salinity of humans is between pure water and sea water:

$$\mu_{\text{sea water}} \underset{\longleftarrow}{<} \mu_{\text{human}} \underset{\longleftarrow}{<} \mu_{\text{fresh water}}$$
$$\text{osmotic pressure}$$

So if we drink fresh water, it diffuses into our organs, which is good. But if we drink sea water, pure water diffuses out of our organs into the sea water, and we dehydrate.

9.3   Chemical Equilibrium

Because Gibbs energy is $G = E - TS + PV = \mu N$, or really $\sum \mu_i N_i$, chemists tend to use G to discuss chemical equilibrium. For isothermal isobaric processes, we have $dG = \mu \, dN$. Now focus on the increase in G for two systems interacting diffusively:

$$dG_1 = \mu_1 \, dN_1 \, ,$$
$$dG_2 = \mu_2 \, dN_2 = -\mu_2 \, dN_1 \, . \qquad (9.25)$$

Note that G is extensive.

> Why? At equilibrium, $G_1 = \mu N_1$ and $G_2 = \mu N_2$, so that $G_1 + G_2 = \mu N = G$. So G scales linearly with the system, which is what extensive variables do.

Because it's extensive, we can write

$$dG = dG_1 + dG_2 = (\mu_1 - \mu_2) \, dN_1 \, , \qquad (9.26)$$

and regardless of whether μ_1 is greater or less than μ_2, this expression will be negative. That is, G tends to decrease, until at equilibrium it must be a minimum. We'll use this in the next discussion.

Direction of a Reaction and Law of Mass Action

Suppose we have molecules A, B, C that can react in either direction:

$$aA + bB \rightleftharpoons cC \, . \qquad (9.27)$$

At some point in time, it's known that the species have chemical potentials μ_A, μ_B, μ_C. We can use these potentials to determine the direction in which the reaction will proceed, by using the fact that G always decreases on the way to equilibrium. So calculate ΔG for each direction of the reaction; the direction for which it's negative will be the direction in which the reaction proceeds. Previously we wrote $dG = \mu \, dN$, but when there is more than one particle species present we should write $dG = \sum \mu_i \, dN_i$, or

$$\Delta G \simeq \sum \mu_i \, \Delta N_i \, . \qquad (9.28)$$

Calculate ΔG for each direction. Remembering the comments in Section 4.1, we must ensure each ΔN_i has the correct sign.

Left to right: When the mixture loses a molecules of A (i.e. $-\Delta N_A = a$) and b molecules of B ($-\Delta N_B = b$), it gains c molecules of C ($\Delta N_C = c$):

$$\Delta N_A = -a, \quad \Delta N_B = -b, \quad \Delta N_C = c. \tag{9.29}$$

When this happens,

$$\Delta G_{L \to R} \simeq -a\mu_A - b\mu_B + c\mu_C. \tag{9.30}$$

Right to left: Now everything is reversed: the mixture gains a molecules of A ($\Delta N_A = a$) and so on. Hence all the signs in the calculation of ΔG are reversed from above:

$$\Delta G_{R \to L} = -\Delta G_{L \to R}. \tag{9.31}$$

Certainly one of either $\Delta G_{L \to R}$ or $\Delta G_{R \to L}$ is negative (unless they are both zero, in which case the reaction has attained equilibrium). This negative one tells us in which direction the reaction goes. You can see why μ is called the chemical potential.

If you remember that Δ always refers to an *increase*, you'll always get the signs right in analyses like the above.

The increases in particle numbers for the left-to-right version of the above reaction, written in (9.29), are called its *stoichiometric coefficients*. Consider a more general reaction with stoichiometric coefficients $b_1, b_2, b_3, \ldots$. As in (9.30), we have

$$\Delta G_{L \to R} \simeq \sum \mu_i b_i. \tag{9.32}$$

But we saw in (9.6) that each chemical potential can be written as

$$\begin{aligned}
\mu_i &= kT \ln \varrho_i + \text{function}_i(T) \\
&\equiv kT \left[\ln \varrho_i - \ln \zeta_i(T) \right],
\end{aligned} \tag{9.33}$$

where ϱ_i are the particle densities and ζ_i are defined for convenience in the following calculation. Then

$$\Delta G_{L \to R} \simeq \sum b_i kT \left[\ln \varrho_i - \ln \zeta_i(T) \right], \tag{9.34}$$

in which case

$$\begin{aligned}
\exp \frac{\Delta G_{L \to R}}{kT} &= \exp \sum b_i \left[\ln \varrho_i - \ln \zeta_i(T) \right] \\
&= \frac{\prod_i \varrho_i^{b_i}}{\prod_i \zeta_i^{b_i}(T)}.
\end{aligned} \tag{9.35}$$

The denominator in this expression is called the *equilibrium constant* for the reaction, which we'll write as $A(T)$:

$$A(T) \equiv \zeta_1^{b_1} \zeta_2^{b_2} \zeta_3^{b_3} \ldots \tag{9.36}$$

Examining (9.35) leads to the following:

$$\prod \varrho_i^{b_i} < A(T) \iff \Delta G_{L\to R} < 0 \iff \text{reaction goes from left to right,}$$

$$\prod \varrho_i^{b_i} > A(T) \iff \Delta G_{L\to R} > 0 \iff \text{reaction goes from right to left,}$$

$$\prod \varrho_i^{b_i} = A(T) \iff \Delta G_{L\to R} = 0 \iff \text{reaction is at equilibrium.} \tag{9.37}$$

In particular, the last line above is known as the *law of mass action*: it tells us the densities of the various species present at equilibrium. In practice the ϱ_i are usually expressed as molar densities (e.g. moles per litre), which is okay because the units will be wrapped up inside $A(T)$.

Example: Writing "mol/ℓ" as M, the equilibrium concentrations in the reaction

$$2A + B \rightleftharpoons 5C + 3D$$
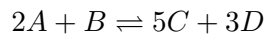
are $\varrho_A = 1$ M, $\varrho_B = 2$ M, $\varrho_C = 3$ M. For the temperature at which the reaction occurs, the equilibrium constant is $100\,\mathrm{M}^5$. What is the concentration of D?

The law of mass action says that at equilibrium,

$$\varrho_A^{-2}\,\varrho_B^{-1}\,\varrho_C^5\,\varrho_D^3 = 100\,\mathrm{M}^5. \tag{9.38}$$

Thus

$$\varrho_D^3 = \varrho_A^2\,\varrho_B\,\varrho_C^{-5}\,100\,\mathrm{M}^5$$

$$= (1\,\mathrm{M})^2\,(2\,\mathrm{M})\,(3\,\mathrm{M})^{-5}\,100\,\mathrm{M}^5 \;= 0.82\,\mathrm{M}^3\,, \tag{9.39}$$

from which it follows that

$$\varrho_D = 0.94\,\mathrm{M}\,. \qquad \underline{\text{Answer}} \tag{9.40}$$

Note that if we had gotten the relative signs wrong in our set of stoichiometric coefficients for the above example, the units wouldn't have worked out right.

10 Fluctuations for a System in Contact with a Reservoir

Statistical methods seemingly can't be used to study small systems such as a single atom. But suppose a small system is in contact with another, larger, system—one so large that its parameters don't change significantly when it interacts with anything else. This larger system is called a *reservoir*, or a *bath*. Since the bath *can* be treated statistically, we can use the fact that the bath and system affect each other, to calculate what happens to the system.

So we ask the question: what fluctuations in a (smaller) system's parameters will occur when it's in equilibrium with a reservoir? We begin with a preliminary treatment before delving into the *Boltzmann Distribution* in Section 10.2. The probability that the system will be found in some configuration is proportional to the number of the accessible states Ω for the system + reservoir in that configuration. Focus on the fluctuations of some parameter $\chi \in \{E, T, P, V, \mu, N\}$. At equilibrium $\chi = \chi_0$, but now a fluctuation $\Delta\chi$ occurs, so

that the new value is $\chi = \chi_0 + \Delta\chi$. We require the probability that the system + reservoir could be in a state with χ instead of χ_0. The probability of a fluctuation $\Delta\chi$ is

$$p(\Delta\chi) \propto \Omega(\chi) = e^{S_{\text{tot}}(\chi)/k}\,, \tag{10.1}$$

where S_{tot} = the sum of the entropies of the system and reservoir. Ignoring all other parameters for conciseness, Taylor's theorem gives

$$S_{\text{tot}}(\chi) = S_{\text{tot}}(\chi_0 + \Delta\chi) = S_{\text{tot}}(\chi_0) + \underbrace{S'_{\text{tot}}(\chi_0)}_{=\,0 \text{ at equilib.}} \Delta\chi + S''_{\text{tot}}(\chi_0)\,\Delta\chi^2/2 + \dots \tag{10.2}$$

So for small fluctuations $\Delta\chi$,

$$p(\Delta\chi) \propto \exp\left(S''_{\text{tot}}(\chi_0)\,\frac{\Delta\chi^2}{2k}\right) \equiv \exp\frac{-\Delta\chi^2}{2\sigma_\chi^2}, \tag{10.3}$$

where σ_χ is a gaussian characteristic spread of the fluctuations. Hence

$$\sigma_\chi^2 = \frac{-k}{S''_{\text{tot}}(\chi_0)}\,. \tag{10.4}$$

Because other variables really are present, we should write this as

$$\sigma_\chi^2 = \frac{-k}{\left.\partial^2 S_{\text{tot}}/\partial\chi^2\right|_{\chi_0}}\,. \tag{10.5}$$

Example. What is σ_E/E for one mole of an ideal monatomic gas?

We require $\partial^2 S_{\text{tot}}/\partial E^2$ at equilibrium. Write the entropy of the system as S and that of the reservoir as S_r, then use conservation of total energy, volume, and particle number to write

$$\mathrm{d}S = \frac{\mathrm{d}E}{T} + \frac{P\,\mathrm{d}V}{T} - \frac{\mu\,\mathrm{d}N}{T}\,,$$

$$\mathrm{d}S_r = \frac{\mathrm{d}E_r}{T_r} + \frac{P_r\,\mathrm{d}V_r}{T_r} - \frac{\mu_r\,\mathrm{d}N_r}{T_r} = \frac{-\mathrm{d}E}{T_r} - \frac{P_r\,\mathrm{d}V}{T_r} + \frac{\mu_r\,\mathrm{d}N}{T_r}. \tag{10.6}$$

Add these to get

$$\mathrm{d}S_{\text{tot}} = \left(\frac{1}{T} - \frac{1}{T_r}\right)\mathrm{d}E + \left(\frac{P}{T} - \frac{P_r}{T_r}\right)\mathrm{d}V - \left(\frac{\mu}{T} - \frac{\mu_r}{T_r}\right)\mathrm{d}N\,. \tag{10.7}$$

Then

$$\left(\frac{\partial S_{\text{tot}}}{\partial E}\right)_{V,N} = \frac{1}{T} - \frac{1}{T_r} \tag{10.8}$$

and T_r is constant (because the reservoir is huge, by definition), so

$$\left(\frac{\partial^2 S_{\text{tot}}}{\partial E^2}\right)_{V,N} = \frac{-1}{T^2}\left(\frac{\partial T}{\partial E}\right)_{V,N}\,. \tag{10.9}$$

Model the system energy's temperature dependence as

$$E = u_0 + \nu N \frac{kT}{2} \,. \tag{10.10}$$

After some algebra,

$$\frac{\sigma_E}{E} = \sqrt{\frac{\nu N \, k^2 T^2}{2 \left(u_0 + \nu \, NkT/2\right)^2}} \,. \tag{10.11}$$

For an ideal gas, $u_0 = 0$, and we get

$$\frac{\sigma_E}{E} = \sqrt{\frac{2}{\nu N}} \,. \tag{10.12}$$

The $1/\sqrt{N}$ is the signature of a relative fluctuation. For one mole of an ideal monatomic gas this is

$$\frac{\sigma_E}{E} = \sqrt{\frac{2}{6^{\underline{23}} \times 3}} \; \simeq \; 10^{-12}. \qquad \underline{\underline{\text{Answer}}} \tag{10.13}$$

10.1 The Concept of the Ensemble

When calculating quantities relating to a system in contact with a bath, it can be helpful to picture a large number of identically prepared systems, each interacting with its own bath, and each in some different (random) stage of its evolution. This imagined set of system–bath pairs is called an *ensemble*. We can treat each system of the ensemble as a fixed point in phase space, with the whole assembly of points comprising the path that a single system would trace out in phase space as it evolved. This idea is called the *ergodic principle*: it suggests that we can convert averages over *time* to averages over *the ensemble*. While historically the ergodic principle has never been completely validated, it is used frequently in statistical mechanics.

Ensembles are classified by the extent of their interaction with a bath:

No interaction: "microcanonical ensemble" (system isolated; its energy conserved)

Thermal/mechanical interaction: "canonical ensemble" (air molecules)

Thermal/mechanical/diffusive interaction: "grand canonical ensemble" (ice crystals interacting with moist air).

10.2 The Boltzmann Distribution

Suppose the interaction increases the system's energy, volume, and particle number by $\Delta E, \Delta V, \Delta N$. The bath's parameters become $E - \Delta E, V - \Delta V, N - \Delta N$. The probability p that the system will have the stated energy, volume, and particle number is proportional to the number of states accessible to the system + bath:

$$p \propto \Omega_s \, \Omega_R \,. \tag{10.14}$$

How to calculate these two numbers of accessible states? The system is probably so simple that the number of states Ω_s accessible to it are easily enumerated. On the other hand,

the bath is so huge that we can't easily enumerate the number of its accessible states Ω_R from first principles; however, we *can* treat it statistically, so can calculate Ω_R from a knowledge of its entropy $S_R = k \ln \Omega_R$. The above probability is

$$p \propto \Omega_s \, e^{S_R/k} \,. \tag{10.15}$$

The First Law for the bath, $E_R = TS_R - PV_R + \mu N_R$, rearranges to give the bath's entropy as

$$S_R = \frac{E_R + PV_R - \mu N_R}{T} = \frac{E - \Delta E + P(V - \Delta V) - \mu(N - \Delta N)}{T} \,, \tag{10.16}$$

so that (10.15) becomes

$$p \propto \Omega_s \, \exp \frac{E - \Delta E + P(V - \Delta V) - \mu(N - \Delta N)}{kT} \,. \tag{10.17}$$

Absorbing E, V, N into the constant of proportionality gives

$$p \propto \Omega_s \, \exp \frac{-\Delta E - P\Delta V + \mu \, \Delta N}{kT} \,. \tag{10.18}$$

We are free to measure $\Delta E, \Delta V, \Delta N$ relative to arbitrary reference levels E_0, V_0, N_0. That is, if the system originally had E_0, V_0, N_0 and now it has $E_s = E_0 + \Delta E$, $V_s = V_0 + \Delta V$, and $N_s = N_0 + \Delta N$, then (10.18) becomes

$$p \propto \Omega_s \, \exp \frac{-(E_s - E_0) - P(V_s - V_0) + \mu \, (N_s - N_0)}{kT}$$

$$\propto \Omega_s \, \exp \frac{-E_s - PV_s + \mu \, N_s}{kT} \,, \tag{10.19}$$

where the system's energy E_s might not *all* have come from an interaction with the bath.

For now we'll restrict ourselves to the case in which no particles are exchanged with the reservoir. The PV_s term is significant only in exceptional circumstances of high pressure, such as in a neutron star. So usually write (10.19) as

$$\boxed{p \propto \Omega_s \, e^{-E_s/(kT)} \,.} \tag{10.20}$$

This is the famous *Boltzmann Distribution*, and is one of the central equations of statistical mechanics.

Example: Suppose a system of hydrogen gas is in contact with a bath at $T = 295 \, \text{K}$. What ratio of the H atoms will be in the 1$^{\text{st}}$ excited state, compared to the ground state? Use $E_0 = -13.6 \, \text{eV}$, $E_1 = -3.4 \, \text{eV}$.

$$\frac{p(E_1)}{p(E_0)} = \frac{\Omega_1 \, e^{-E_1/(kT)}}{\Omega_0 \, e^{-E_0/(kT)}} = \frac{\Omega_1}{\Omega_0} \, e^{-(E_1 - E_0)/(kT)} \,. \tag{10.21}$$

In the ground state, the H quantum numbers are $(n \, l \, m) = (1 \, 0 \, 0)$, with two possible electron spins, so $\Omega_0 = 2$. In the 1$^{\text{st}}$ excited state the quantum numbers can be $(2 \, 0 \, 0), (2 \, 1 \, 1), (2 \, 1 \, 0), (2 \, 1 - 1)$, each with two spins, so $\Omega_1 = 8$. We have

$$kT \simeq \frac{1.381 \overset{-23}{=\!=} \times 295}{1.602 \overset{-19}{=\!=}} \, \text{eV} \simeq 0.0254 \, \text{eV} \,. \tag{10.22}$$

Then

$$\frac{p(E_1)}{p(E_0)} = \frac{8}{2} \exp \frac{-10.2\,\text{eV}}{0.0254\,\text{eV}} \simeq 2 \times 10^{-174}. \qquad \underline{\text{Answer}} \qquad (10.23)$$

Clearly, most of the atoms are unexcited.

> It's useful to note that at room temperature, $kT \simeq 1/40$ eV. Remember this number: it forms a good rule of thumb to help you determine quickly whether much of a system is excited or not. In the above example, the energy "distance" to the first excited state is 10.2 eV, and this is so much larger than 1/40 eV that we can see immediately that there will be almost no excitation. The typical energy supplied by interactions with the bath, kT, is just too small to excite many atoms—although, fluctuations being what they are, *some* few atoms will be excited.

In general, we can define the *excitation temperature* of a system as T_e such that

$$\frac{E_1 - E_0}{kT_e} \equiv 1\,. \qquad (10.24)$$

The excitation temperature gives an indication of the temperature at which an appreciable number of particles begin to occupy the 1st excited state.

10.3 Diatomic Gases and Heat Capacity

In (6.11) we wrote $\gamma = (\nu + 2)/\nu$ in the context of heat capacity. At room temperature, γ is measured to have the following values:

molecule:	HCl	NO	Cl_2	Br_2	I_2
γ:	1.41	1.40	1.36	1.32	1.30

These values begin at 7/5, so diatomic gases such as HCl and NO seem to be rigid rotors with 5 degrees of freedom: presumably 3 translational and 2 rotational. For a non-rigid rotor, we expect vibration to contribute 2 more degrees of freedom, making $\gamma = 9/7 \simeq 1.29$. So iodine seems to be a non-rigid rotor at room temperature, with chlorine and bromine somewhere in between.

It turns out that the heat capacity for any particular gas varies with temperature as the degrees of freedom go from purely translational to translational + rotational, to translational + rotational + vibrational. Let's investigate this using quantum mechanics.

10.3.1 Rotation

A rigid diatomic molecule can only have rotational energy levels of

$$E_\ell = \frac{\ell(\ell+1)\hbar^2}{2I}\,, \quad \text{where } \ell = 0, 1, 2, \ldots \qquad (10.25)$$

In a gas at temperature T, we have relative populations

$$\frac{N(E_\ell)}{N(E_0)} = \frac{\Omega_\ell\, e^{-E_\ell/(kT)}}{\Omega_0\, e^{-E_0/(kT)}}\,, \qquad (10.26)$$

where $\Omega_\ell = 2\ell + 1$ is the usual quantum mechanical degeneracy associated with the set $-\ell, \ldots, \ell$. Thus

$$\frac{N(E_\ell)}{N(E_0)} = (2\ell + 1)\, \exp \frac{-\ell(\ell+1)\frac{\hbar^2}{2I}}{kT} \,. \tag{10.27}$$

Define the characteristic temperature of (the onset of) rotation as T_R, where

$$kT_R \equiv \frac{\hbar^2}{2I} \,. \tag{10.28}$$

Then

$$\frac{N(E_\ell)}{N(E_0)} = (2\ell + 1)\, e^{-\ell(\ell+1)T_R/T} \,. \tag{10.29}$$

If $T \ll T_R$ then $N(E_\ell)/N(E_0) \approx 0$, and the rotational states are "frozen out". The rotational energy level spacing is large compared to the ambient supply of thermal energy kT, which is why rotational states can't be activated.

If $T \gg T_R$ then $N(E_\ell)/N(E_0) > 0$ and the rotational energy states are well populated. The rotational energy level spacing is now small compared to kT. With so many rotational states able to be accessed, a gas of such molecules well and truly has rotational degrees of freedom, and can be treated using the Equipartition Theorem.

Rotation about a non-internuclear axis   Consider a classical picture of a rigid rotor formed from two masses lying along the x axis, that spin around the z axis about their centre of mass, which lies at the origin. The masses forming the rotor are m_1, distance r_1 from the origin, and m_2, distance r_2 from the origin. The two masses are a distance D apart. What is T_R? We need the moment of inertia I_z for rotation about the z axis:

$$I_z = m_1 r_1^2 + m_2 r_2^2 \,. \tag{10.30}$$

But if the centre of mass is at the origin, then $-m_1 r_1 + m_2 r_2 = 0$. We can use this to show that

$$I_z = \mu D^2 \,, \quad \text{where} \quad \frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \,. \tag{10.31}$$

μ is called the *reduced mass* of the system.

Example: Calculate the characteristic temperature of rotation for CO, given that the C and O atoms are a distance of $0.112\,\text{nm}$ apart.

The masses of an atom of carbon and an atom of oxygen are $12\,\text{g}/N_A$ and $16\,\text{g}/N_A$ respectively, and $D = 0.112\,\text{nm}$. Thus

$$\begin{aligned}
T_R &= \frac{\hbar^2}{2I_z k} = \frac{\hbar^2}{2\mu D^2 k} = \frac{\left[6.626^{\underline{-34}}/(2\pi)\right]^2 \times 6.022^{\underline{23}} \left(\frac{1}{0.012} + \frac{1}{0.016}\right)}{2 \times \left(0.112^{\underline{-9}}\right)^2 \times 1.38^{\underline{-23}}} \,\text{K} \\
&= 2.8\,\text{K} \qquad \underline{\text{Answer}}
\end{aligned} \tag{10.32}$$

So at room temperature a gas of CO molecules has many rotational states occupied, and the Equipartition Theorem can be applied to it.

Rotation about the internuclear axis For the same masses above, we can use a classical picture to calculate the moment of inertia I_x for rotation about the x axis, i.e. about the line joining the masses. We find that $I_x \ll I_z$. For CO, a similar calculation to the one above gives the characteristic temperature for the onset of this type of rotation as $\approx 100{,}000\,\text{K}$; hence such rotation is frozen out at room temperature. In fact, quantum mechanically, there can be no rotation at all about the x axis, so this rotation is frozen out at *all* temperatures.

10.3.2 Vibration

A harmonic oscillator of frequency f can only have vibrational energies

$$E_n = (n + {}^1\!/_2)\, hf\,, \quad \text{where } n = 0, 1, 2, \dots \tag{10.33}$$

Relative populations are

$$\frac{N(E_n)}{N(E_0)} = \frac{\Omega_n \, e^{-(n+1/2)hf/(kT)}}{\Omega_0 \, e^{\frac{-1}{2}hf/(kT)}} = e^{-n\,hf/(kT)}\,, \tag{10.34}$$

because $\Omega_n = 1$ for all n. Define the characteristic temperature of (the onset of) vibration as T_V, where

$$kT_V \equiv hf\,. \tag{10.35}$$

Then

$$\frac{N(E_n)}{N(E_0)} = e^{-n\,T_V/T}\,. \tag{10.36}$$

If $T \ll T_V$ then $N(E_n)/N(E_0) \approx 0$, and the vibrational levels are frozen out. The vibrational energy level spacing is large compared to kT.

If $T \gg T_V$ then $N(E_n)/N(E_0) > 0$ and the vibrational energy levels are well populated. The vibrational energy level spacing is now small compared to kT. With the vibrational levels well occupied, a gas of such molecules has vibrational degrees of freedom, and can be treated using the Equipartition Theorem. Typical values of T_V for light molecules are several thousand kelvins, so at room temperature and beyond, these molecules don't vibrate.

10.4 Equipartition for a System Touching a Bath

The Equipartition Theorem was stated originally in Section 5.1 for what amounted to an isolated system. We saw there that each degree of freedom contributes ${}^1\!/_2\,kT$ to the total energy. What is its equivalent for a system in contact with a bath? Fluctuations now allow the system to have a range of energies. For such a case, let's calculate the *mean* energy contributed to the system by each degree of freedom.

The mean thermal energy associated with any particular coordinate u is

$$\langle E_u \rangle = \int_0^\infty E_u \, p(E_u)\, \mathrm{d}E_u\,, \tag{10.37}$$

where

$$p(E_u)\, \mathrm{d}E_u = \text{prob. that the system has energy in } E_u \to E_u + \mathrm{d}E_u$$
$$= (\text{prob. system in a state with } E_u) \times \text{no. of states in that interval.} \tag{10.38}$$

The probability of being in a state with energy E_u is $\propto e^{-E_u/(kT)}$. The number of states in the energy interval dE_u equals the number of states in the coordinate interval du. In Section 3.1, we showed that this number is proportional to the length du of that interval. Hence

$$p(E_u)\,dE_u = Ae^{-E_u/(kT)}\,du\,, \quad \text{for some normalisation } A. \qquad (10.39)$$

As before, we'll just consider quadratic energy dependence: $E_u = bu^2$ for some b. Calculate A by demanding that $\int_0^\infty p(E_u)\,dE_u = 1$. (Note that it suffices to consider positive u only, since E_u is even in u. In fact, if we use the whole range of u we will get a different normalisation, but the end result for $\langle E_u \rangle$ will be the same.) Using the gaussian integral (2.27) it's easy to show that

$$A = 2\sqrt{\frac{b}{\pi kT}}\,. \qquad (10.40)$$

Now we can evaluate (10.37):

$$\langle E_u \rangle = \int_0^\infty bu^2 A e^{-bu^2/(kT)}\,du$$

$$= bA \int_0^\infty u\,u\,e^{-bu^2/(kT)}\,du\,. \qquad (10.41)$$

This last equation can be integrated by parts (the parts are u and $u\,e^{-bu^2/(kT)}$) to give

$$\boxed{\langle E_u \rangle = {}^1\!/_2\,kT\,.} \qquad (10.42)$$

[For an alternative way to integrate (10.41), see (12.5) and (12.6).] So this is the generalisation of the Equipartition Theorem to the case of a non-isolated system. Each degree of freedom now contributes an *average* value of ${}^1\!/_2\,kT$ to the thermal energy.

11   Entropy of a System Touching a Bath

The central postulate of statistical mechanics is that the accessible states of an isolated system are all equally likely. Calling the number of these Ω, we defined the thermodynamic entropy $S \equiv k\ln\Omega$ in Section 5. From a purely mathematical viewpoint, we also had the statistical entropy $\sigma \equiv \ln\Omega$, which just omits Boltzmann's constant.

A system $\mathcal{S}$ in contact with a bath is no longer isolated, so its accessible states are in general not equally likely. Can we still find an expression for its entropy? Suppose its states are labelled $|1\rangle, |2\rangle, \ldots, |M\rangle$, where M might be infinite. If the probabilities p_i of being found in a state $|i\rangle$ are all equal, we can set $\Omega = M$ and write $\sigma = \ln M$. When the probabilities are not all equal, we form an (isolated) ensemble of a huge number N of distinguishable copies of $\mathcal{S}$. As $N \to \infty$, the number of copies of $\mathcal{S}$ found in state $|i\rangle$ is $n_i = Np_i$.

Now, how to count the accessible states? This number will be hugely dominated by the number pertaining to equilibrium, being the number of ways that we can arrange

n_1 systems in $|1\rangle$, n_2 systems in $|2\rangle$, etc. Because of the incredibly highly peaked nature of Ω, we can set Ω equal to this equilibrium number and not include all the other ways, such as $n_1 + n_2$ copies in $|1\rangle$, none in $|2\rangle$, n_3 in $|3\rangle$, etc. We calculated this number for $M = 2$ states in Section 2 by considering the binomial distribution. There we found:

$$\text{Number of ways of putting } n_1 \text{ systems in } |1\rangle \text{ and } n_2 \text{ systems in } |2\rangle = \frac{N!}{n_1!\, n_2!}, \quad (11.1)$$

where $N = n_1 + n_2$. The same reasoning applies to the case of more than 2 states. Label N distinguishable particles $1, 2, \ldots, N$ and allocate each to one of M bins; then apply the approach outlined in (2.1). Each combination will occur $n_1!\, n_2! \, \ldots \, n_M!$ times, so the total number of *permutations*, $N!$, over-counts by this factor. Thus the required total number of combinations must be

$$\frac{N!}{n_1!\, n_2! \, \ldots \, n_M!} \, . \quad (11.2)$$

The actual probability distribution that extends the binomial case is called the *multinomial distribution*. That is, if the chance of a particular particle being allocated to bin i is p_i, then what is the chance $P(n_1, n_2, \ldots, n_M)$ of finding n_i particles, without regard for order, in bin i for some given $n_1, n_2, \ldots, n_M$? Each combination described in the previous paragraph occurs with probability $p_1^{n_1}\, p_2^{n_2} \ldots p_M^{n_M}$, so

$$P(n_1, n_2, \ldots, n_M) = \frac{N!}{n_1!\, n_2! \, \ldots \, n_M!}\, p_1^{n_1}\, p_2^{n_2} \ldots p_M^{n_M} \, . \quad (11.3)$$

This is useful to know, but we only need the total number of combinations $N!/(n_1!\, n_2! \, \ldots \, n_M!)$.

The number of combinations in (11.2) pertains to the whole ensemble so, to the accuracy mentioned above, it must equal Ω^N, since the number of states is multiplicative just as entropy is additive. In that case the statistical entropy is

$$\sigma = \ln \Omega = \frac{1}{N} \left[\ln N! - \sum_{i=1}^{M} \ln n_i! \right]$$

$$\simeq \frac{1}{N} \left\{ (N + 1/2) \ln N - \cancel{N} + \underbrace{\ln \sqrt{2\pi}}_{\equiv \ell} - \sum_i \left[(n_i + 1/2) \ln n_i - \cancel{n_i} + \ln \sqrt{2\pi} \right] \right\}$$

$$= \left(1 + \frac{1}{2N} \right) \ln N + \frac{\ell}{N} - \sum_i \left(p_i + \frac{1}{2N} \right) \ln(Np_i) - \frac{M\ell}{N} \, . \quad (11.4)$$

(Note that this expression is not overly changed if $M \to \infty$, since the states can't *all* have large occupation numbers. Stirling's approximation only applies to those with large n_i, and we'll ignore the rest because they are not well populated.) Writing $\ln(Np_i) = \ln N + \ln p_i$, regrouping and cancelling terms gives

$$\sigma = \frac{1 - M}{2N} \ln N + (1 - M)\frac{\ell}{N} - \sum_i p_i \ln p_i - \frac{1}{2N} \sum_i \ln p_i$$

$$\to -\sum_i p_i \ln p_i \quad \text{as } N \to \infty \, . \quad (11.5)$$

This last expression is important and famous: it's the statistical entropy of a single system, and is worth rewriting in a box:

$$\boxed{\sigma = -\sum_i p_i \ln p_i \, .}$$ (11.6)

(Thus the thermodynamic entropy is $S = -k \sum_i p_i \ln p_i$.) What does (11.6) give for the entropy of an isolated system? The fundamental postulate of statistical mechanics (on page 7) says that all states are equally possible for an isolated system in equilibrium. If the M states are all equally possible, then $p_i = 1/M$ for all i. Then (11.6) becomes

$$\sigma = -\sum_{i=1}^{M} \frac{1}{M} \ln \frac{1}{M} = -\ln \frac{1}{M} = \ln M \, .$$ (11.7)

This is just what we expect from first principles, because in such a case of M equiprobable states we'd write $\Omega = M$, so that $\sigma = \ln \Omega = \ln M$. So the above ensemble analysis is consistent with the fundamental definition of entropy for an isolated system that we have been using throughout this course.

11.1   A Brief Information Primer

The expression for entropy that we have just found was known by physicists long before it was rediscovered in a new context by Claude Shannon in the 1940s, who pioneered the field of information theory.

Central to information theory is the idea of the probability p_i that the next symbol in a sequence being transmitted will be symbol i. If $p_i \approx 0$ then symbol i is rare, so we'll be surprised to see it; in that sense, it has a high "information-transmitting potential". If $p_i \approx 1$, symbol i is common, so we are not surprised to see it: it can't have much information-transmitting potential if it was expected anyway. Define the "surprise at seeing symbol i" to be $-\log_b p_i$. The number b is usually set equal to 2 by information theorists, as it relates to storing information in a binary way using e.g. on/off settings of a switch. We'll leave it as a general number b here. In fact, for any a, b, c,

$$\log_a b = \frac{\log_c b}{\log_c a} \, ,$$ (11.8)

so the logarithm to any base is just equal to a constant times the natural logarithm that we have been using exclusively in statistical mechanics.

As the probability p_i that the next symbol will be i ranges from 0 to 1, the surprise $-\log_b p_i$ associated with seeing it ranges from ∞ to 0, consistent with the above discussion. Our average surprise on seeing the next symbol (i.e. averaged over all symbols of the alphabet being used) is

$$\text{average surprise} \equiv \langle -\log_b p \rangle = -\sum_i p_i \log_b p_i = \frac{-1}{\ln b} \sum p_i \ln p_i \, .$$ (11.9)

Note that we could have used another expression, such as $1/p_i - 1$, for the surprise at seeing symbol i, that would also range from ∞ to 0. However, Shannon's original analysis showed that (11.9) is the only possible expression for what we have called the average surprise that is consistent with a set of requirements that he laid down for such a quantity.

The average surprise is called the *Shannon entropy* of the alphabet, and usually b is set equal to 2; in other words, Shannon replaced the Boltzmann constant of our expression for entropy by $1/\ln 2$ in his context. This average surprise can be shown to be maximal when all the p_i are equal (we'll do so below for a 2-symbol alphabet). In that sense, a high entropy means each letter is being well used. Such an alphabet has a high information-transmitting potential:

$$\text{information-transmitting potential} \equiv \text{Shannon entropy}$$
$$= \text{average surprise when } b = 2. \qquad (11.10)$$

Example: Describe the information-transmitting potential of an alphabet of two symbols; i.e., when two symbols are available to be transmitted. Symbol 1 appears with probability p_1 and symbol 2 appears with probability $p_2 = 1 - p_1$.

There is only one free variable: choose it to be p_1. Also, to show that base 2 isn't actually necessary to this discussion, we won't set $b = 2$. Then (11.9) produces

$$I(p_1) \equiv \text{inf.-trans. potential} = \frac{-1}{\ln b}\big[p_1 \ln p_1 + (1 - p_1)\ln(1 - p_1)\big]. \qquad (11.11)$$

We leave it to you to show that $I(0) = I(1) = 0$. Also, $I'(p_1) = \log_b(1/p_1 - 1)$, which is zero when $p_1 = 1/2$. This leads to a graph of $I(p_1)$ versus p_1 that is everywhere concave down, rising from zero at the endpoints to a maximum at the midpoint of $p_1 = p_2 = 1/2$, and symmetrical about that midpoint. So the information-transmitting potential (Shannon entropy) of this small alphabet is maximal when each symbol is equally allowed to appear—which sounds reasonable—and zero when only one symbol is allowed to appear. Again this is quite reasonable; after all, if sentences using the alphabet were primarily composed of just one of the symbols, then the alphabet wouldn't be of much use.

It's not hard to show—using the method of Lagrange multipliers—that the same conclusion is true for an alphabet of any length. The information-transmitting potential of an alphabet of M letters is again maximal when each letter is equally allowed to appear, and has the value $\langle -\log_b 1/M \rangle = \log_b M$ where b is arbitrary.

Example: We can calculate the information-transmitting potential, or entropy, of the English language as follows. Take a typical book that represents the language as it's normally used. Count the frequencies of the letters and punctuation and use these to compile a set of probabilities for those characters. For example, when the letters "a" to "z" and spaces are counted, a set of probabilities $p_1, \ldots, p_{27}$ is produced. (These are not all equal; for instance, the probability of an "e" is comparatively high, and so on.) One such set of typical probabilities gives $-\sum p_i \ln p_i \simeq 2.83$. Thus the information-transmitting potential of English is about 2.83 (using the natural logarithm). Now suppose that English were to be replaced by a new alphabet in which each letter was equally able to appear. How many letters would be required to match the information-transmitting potential of English? Call this number of letters M. Then the fact that all of the new p_i are equal implies that the new alphabet's information-transmitting potential is $\ln M$. This is required to equal 2.83, so $M = e^{2.83} \simeq 17$. That is, the new alphabet would need just 17 letters.

That English could get by with 17 letters does not imply that it *should* have 17 letters. Redundancy in information flow can be useful because it makes the job of processing that information easier for our brains—as well as giving us time to savour what is being

imparted. If English were to be pared down to become entropically efficient (but perhaps lifeless), our level of concentration would have to increase to ensure we didn't miss any details of its tight transmission. That would tend to introduce errors of its own.

Additionally, if our task were to trim our alphabet to a set that fulfils a computer's expectations of efficiency, then where would such trimming end? Would each letter be streamlined to minimise its use of ink and complexity of shape, with the final result being that calligraphy is reduced to nothing more than strings of agitated commas? No, and for the same reason, singers don't replace a song's repeated verse with the word "Ditto". Nor do portrait painters depict only one eye of their subject, arguing that the other is more or less a mirror image and so needn't be drawn.

Entropy and Information

The entropy, or information-transmitting potential, of an alphabet is usually just called its "information" by information theorists, who know what they are doing. However, we shouldn't be misled by this word into thinking that the entropy $-\sum p_i \ln p_i$ is somehow giving the information content of whatever was sampled to give the set of p_i. The concept of information as it's usually understood (as opposed to information-transmitting potential) is not quantifiable, and certainly cannot easily be related to entropy. For example, the entropy, or information-transmitting potential, that we calculated above for English has nothing to do with whatever information might be contained in the sample text that was used to compile the set of probabilities.

To see further why entropy and information content cannot simply be equated, consider a monkey who types a book using a standard 26-letter typewriter. Each letter will probably appear about $1/26$ of the time, so that all the p_i equal $1/26$. Now suppose I write a book. I use a new language in which each word has 26 letters, with each letter from "a" to "z" appearing exactly once. (There are 26! possible words in such a language, more than enough for the job.) Each letter appears $1/26$ of the time in my book too, so again all the p_i equal $1/26$. Yet the monkey's book almost certainly carries no information, whereas presumably my book has a lot of information.

So a book's information content is not simply related to the set of p_i; nor is it easily related to various correlations of letters, although these would have to be taken into account in any deeper analysis. The deepest analysis would completely classify all correlations to such an extent that we would simply end up reading the book; nevertheless, we could still only make a value judgement on how much information it contains. Information theory is not about such things. Rather, it's concerned with information-transmitting potential. And certainly the two *alphabets* of the two languages used by the monkey and myself *do* have high information-transmitting potentials.

11.2   The Brandeis Dice

The following question was made famous by the statistical physicist E.T. Jaynes in his 1962 lectures at Brandeis University.

A (possibly biased) die is thrown many times, and the results are summarised in a single statement: the mean number showing on the top face is 5. What is the best estimate of the probability of getting each of the numbers 1 to 6 on the next throw? The mean of

the numbers that land face up on an unbiased die will be 3.5, so we do know that the probabilities for each of the numbers 1 to 6 cannot all be equal. That's all we can say with certainty, but we *can* make an educated guess as to what the required probabilities might be.

Set p_i to be the probability of face i landing up. Jaynes defined the sought-after best estimate of this probability to be the "blandest" probability function consistent with the constraints of $\sum_{i=1}^{6} i\, p_i = 5$ and $\sum_{i=1}^{6} p_i = 1$. Why? Because we hardly expect it to be otherwise; yes, the function might have an interesting peak: the die might have a 5 on each of its faces so that $p_5 = 1$ and all the other $p_i = 0$, but this is unlikely. (We can still model such a case as the die having 1 to 6 on its faces, with an extreme—and interesting!—bias such that only the 5 face ever lands up.)

Suppose we construct lots of estimates of this set of six probabilities by having a team of monkeys repeatedly construct "three-dimensional metallic" bar graphs by dropping a huge number N of coins into six vertical slots. Jaynes' approach was thus to choose the most common distribution of coins that resulted. That is, if the monkeys drop n_i coins into the ith slot ($i = 1, \ldots, 6$), then we wish to maximise $\Omega(p_1, \ldots, p_6)$, the number of ways of obtaining the set $n_1, \ldots, n_6$, where $p_i = n_i/N$.

Suppose that in general there are M slots ($M = 6$ for a die). Then

$$\Omega = \frac{N!}{n_1!\, n_2!\, \ldots\, n_M!}\,. \tag{11.12}$$

But we've already seen this in (11.2), and we know that when $N \gg M$, maximising Ω is equivalent to maximising $-\sum_{i=1}^{M} p_i \ln p_i$. Jaynes made this an entry point for a new approach to statistical mechanics, one that gave pre-eminence to the entropy $-\sum p_i \ln p_i$.

Let's generalise the die further, so that the number on face i is E_i. If the average number thrown is E, what are all of the p_i? We need to maximise $-\sum p_i \ln p_i$ subject to

$$\sum_{i=1}^{M} p_i E_i = E\,, \qquad \sum_{i=1}^{M} p_i = 1\,. \tag{11.13}$$

Extremising an expression subject to constraints is usually done using Lagrange multipliers. These multipliers are unknowns that are introduced (one for each constraint, and called α and β in the next few lines), such that the following holds for each variable p_i:

$$\frac{\partial}{\partial p_i}\left(-\sum p_i \ln p_i\right) = \alpha\, \frac{\partial}{\partial p_i} \sum p_i E_i + \beta\, \frac{\partial}{\partial p_i} \sum p_i\,. \tag{11.14}$$

(The method of Lagrange multipliers is not meant to be obvious, but proving why it works is left for a maths course.) Doing the partial derivatives gives

$$-\ln p_i - 1 = \alpha E_i + \beta \quad \text{for all } i\,, \tag{11.15}$$

so that

$$p_i = \underbrace{e^{-1-\beta}}_{\text{normalisation}} e^{-\alpha E_i} = \frac{e^{-\alpha E_i}}{\sum_i e^{-\alpha E_i}} \equiv \frac{e^{-\alpha E_i}}{Z}\,, \tag{11.16}$$

where the (reciprocal of the) normalisation is $Z \equiv \sum_i e^{-\alpha E_i}$, a useful quantity in statistical mechanics known as the *partition function*, that we'll meet again in Section 15.

For the case of the die with $E = 5$ and $M = 6$ that we began with, $E_i = i$, so (11.16) gives

$$p_1 = \frac{e^{-\alpha}}{Z}, \quad p_2 = \frac{e^{-2\alpha}}{Z}, \quad \ldots, \quad p_6 = \frac{e^{-6\alpha}}{Z}. \tag{11.17}$$

To find α, substitute the p_i into (11.13) to yield a fifth-order polynomial whose roots must be found. When this is done, we obtain

$$p_1 \simeq 0.02, \ p_2 \simeq 0.04, \ p_3 \simeq 0.07, \ p_4 \simeq 0.14, \ p_5 \simeq 0.25, \ p_6 \simeq 0.48. \tag{11.18}$$

As expected, the probabilities are larger around the number 5.

What about the case of $E = 3.5$? This is the mean for an unbiased die. When we apply Jaynes' procedure as above, we do indeed find that all of the $p_i = 1/6$, as expected. It should be no surprise to find that the blandest die that gives $E = 3.5$ is an unbiased one.

Jaynes' "Brandeis dice" automatically generate the Boltzmann distribution, although of course as presented in this short section, the notion of a temperature still needs to be introduced. Nonetheless, his ideas have proved to be extremely fruitful in statistical mechanics.

12 Distribution of Motions of Gas Particles

When a gas is in contact with a heat bath, its particles will have a distribution of velocities governed by the Boltzmann distribution. We ask two questions:

(a) How many particles will be found in the range of velocities from $\boldsymbol{v}$ to $\boldsymbol{v} + \mathrm{d}\boldsymbol{v}$? Call this infinitesimal number $N(\boldsymbol{v})\,\mathrm{d}^3 v$, where $\mathrm{d}^3 v \equiv \mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z$. The function $N(\boldsymbol{v})$ is the *Maxwell velocity distribution*.

(b) How many particles will be found in the range of speeds from v to $v + \mathrm{d}v$? Call this infinitesimal number $N(v)\,\mathrm{d}v$. The function $N(v)$ is the *Maxwell speed distribution*.

In most situations, a gas can be considered to be in contact with a heat bath. For example, the molecules of the air in the lecture theatre follow a Maxwell distribution. We can calculate these distributions as follows.

12.1 The Maxwell Velocity Distribution

Suppose there are N_{tot} gas particles each with mass m, and that they are distinguishable, so that we can consider the probability that any particular one of them will be found in the velocity interval $\boldsymbol{v}$ to $\boldsymbol{v} + \mathrm{d}\boldsymbol{v}$. This is

$$\frac{N(\boldsymbol{v})\,\mathrm{d}^3 v}{N_{\text{tot}}} = (\text{prob. particle in } v_x \to v_x + \mathrm{d}v_x) \times (\text{ditto } v_y) \times (\text{ditto } v_z)$$

$$= \underbrace{(\text{prob. particle in state with } E = 1/2\,mv_x^2 + \cdots + 1/2\,mv_z^2 = 1/2\,mv^2)}_{\propto \exp \frac{-mv^2}{2kT}}$$

$$\times \underbrace{(\text{no. of states in } E \to E + \mathrm{d}E)}_{= \mathrm{d}\Omega_{\text{tot}} = g(E)\,\mathrm{d}E}. \tag{12.1}$$

We did some density-of-states calculations in Section 3.1, but will rerun the analysis here to show an alternative approach. In Section 3.1 we wrote Ω_{tot} for the number of states with energies $0 \to E$. We calculated Ω_{tot} in (3.9), then differentiated it in (3.10) to get the density of states $g(E)$. But as we wrote in (3.4), $\mathrm{d}\Omega_{\text{tot}} = g(E)\,\mathrm{d}E$. So instead of calculating Ω_{tot} we might choose to focus on $\mathrm{d}\Omega_{\text{tot}}$, the number of states in the range $E \to E + \mathrm{d}E$, as we did in (3.5):

$$\mathrm{d}\Omega_{\text{tot}} = (\text{no. of ``quant. mechanical cells'' for } x \text{ coord.}) \times (\text{ditto } y) \times (\text{ditto } z)$$

$$= \frac{[x][p_x]}{h} \cdots \frac{[z][p_z]}{h} = \frac{1}{h^3} \underbrace{[x]\ldots[z]}_{= V_{\text{space}}} \underbrace{[p_x]\ldots[p_z]}_{= m\,\mathrm{d}v_x \,\ldots\, m\,\mathrm{d}v_z \,\propto\, \mathrm{d}^3 v} . \tag{12.2}$$

Then (12.1) becomes

$$N(\boldsymbol{v})\,\mathrm{d}^3 v \propto e^{\frac{-mv^2}{2kT}}\,\mathrm{d}^3 v\,, \quad \text{or} \ \ N(\boldsymbol{v}) = C\,e^{\frac{-mv^2}{2kT}} \tag{12.3}$$

with normalisation C. Determine C by counting the particles:

$$N_{\text{tot}} = \int\limits_{\text{all velocities}} N(\boldsymbol{v})\,\mathrm{d}^3 v = C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-mv^2}{2kT}} \underbrace{\mathrm{d}^3 v}_{\text{convert to polar: } = v^2 \sin\theta\,\mathrm{d}v\,\mathrm{d}\theta\,\mathrm{d}\phi}$$

$$= C \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta\,\sin\theta \int_0^{\infty} \mathrm{d}v\,v^2 e^{\frac{-mv^2}{2kT}}$$

$$= C \cdot 2\pi \cdot 2 \cdot \int_0^{\infty} \mathrm{d}v\,v^2 e^{\frac{-mv^2}{2kT}} . \tag{12.4}$$

This integral was evaluated in Section 10.4. Alternatively, do it by differentiating

$$\int_0^{\infty} e^{-av^2}\,\mathrm{d}v = \frac{1}{2}\sqrt{\frac{\pi}{a}} \tag{12.5}$$

with respect to a (this is called "differentiation under the integral sign") to get

$$\int_0^{\infty} v^2 e^{-av^2}\,\mathrm{d}v = \frac{\sqrt{\pi}}{4}\,a^{-3/2}. \tag{12.6}$$

Setting $a = \frac{m}{2kT}$ gives the integral. We can then write C in terms of N_{tot}, so that (12.3) becomes a gaussian:

$$\boxed{N(\boldsymbol{v}) = N_{\text{tot}} \left(\frac{m}{2\pi kT} \right)^{3/2} e^{\frac{-mv^2}{2kT}} .} \tag{12.7}$$

Is this reasonable? Imagine drawing a bar graph of the numbers of molecules in a room versus the x components of their velocities. In a first simplistic analysis, divide the molecules into two roughly defined sets: half are moving up/down and half are moving left/right. The half that are moving up/down all have $v_x \approx 0$, so draw a bar of height $1/2\,N_{\text{tot}}$ at $v_x = 0$. For the molecules moving left/right, half move left and half move right, so draw two bars of height $1/4\,N_{\text{tot}}$, at equal distances somewhere to the left and right of $v_x = 0$. We see a symmetrical function that peaks at $v_x = 0$ beginning to take shape.

The width of the gaussian in (12.7) (i.e. its standard deviation) is $\sqrt{kT/m}$. As might be expected, the distribution is broadened by higher temperatures and less massive gas particles.

12.2 The Maxwell Speed Distribution

We are usually only interested in the speeds of the particles, not their directions of motion. What results is the Maxwell speed distribution.

Let $N(v)\,\mathrm{d}v$ be the infinitesimal number of particles found in the range of speeds from v to $v + \mathrm{d}v$:

$$
N(v)\,\mathrm{d}v = \iint_{\text{all directions}} N(\boldsymbol{v})\,\mathrm{d}^3v = \iint_{\text{all directions}} N_{\text{tot}}\left(\frac{m}{2\pi kT}\right)^{3/2} e^{\frac{-mv^2}{2kT}}\,\mathrm{d}^3v
$$

$$
= N_{\text{tot}}\left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^{2\pi}\mathrm{d}\phi \int_0^{\pi}\mathrm{d}\theta\,\sin\theta\,v^2 e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v
$$

$$
= N_{\text{tot}}\sqrt{\frac{2}{\pi}}\left(\frac{m}{kT}\right)^{3/2} v^2 e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v\,. \tag{12.8}
$$

Thus we arrive at

$$
\boxed{N(v) = N_{\text{tot}}\sqrt{\frac{2}{\pi}}\left(\frac{m}{kT}\right)^{3/2} v^2 e^{\frac{-mv^2}{2kT}}\,.} \tag{12.9}
$$

(Remember that speed is always non-negative, so $v \geqslant 0$.) Compare this with the velocity distribution (12.7): aside from the different normalisation, the speed distribution has an extra factor of v^2, which pushes its peak out to some value of speed greater than zero. We'll determine this value soon.

Alternative Derivation of the Speed Distribution

The speed distribution is sometimes calculated slightly differently to what we have done. We derived it by summing over all directions using the velocity distribution. But we could have avoided reference to the velocity distribution as follows. We have

$$
\frac{N(v)\,\mathrm{d}v}{N_{\text{tot}}} = \text{prob. for particle to have speed in } v \to v + \mathrm{d}v
$$

$$
= \underbrace{\left(\text{prob. particle in energy state } \tfrac{1}{2}mv^2\right)}_{\propto\, \exp\frac{-mv^2}{2kT}} \times \underbrace{\left(\text{no. of states in } v \to v + \mathrm{d}v\right)}_{= \,\mathrm{d}\Omega_{\text{tot}} = g(v)\,\mathrm{d}v = g(E)\,\mathrm{d}E}. \tag{12.10}
$$

Now realise that $\mathrm{d}\Omega_{\text{tot}}$ is the number of states in a shell of infinitesimal thickness in momentum space at energy $E = \tfrac{1}{2}mv^2$:

$$
\mathrm{d}\Omega_{\text{tot}} = (\text{no. of "quant. mechanical cells" for } x \text{ coord.}) \times (\text{ditto } y) \times (\text{ditto } z)
$$

$$
= \frac{[x][p_x]}{h}\cdots\frac{[z][p_z]}{h} = \frac{1}{h^3}\underbrace{[x]\ldots[z]}_{=\,V_{\text{space}}}\underbrace{[p_x]\ldots[p_z]}_{\text{volume of shell in 3-D momentum space}}
$$

$$
\propto \text{volume of shell of radius } v \to v + \mathrm{d}v
$$

$$
\propto v^2\,\mathrm{d}v\,. \tag{12.11}
$$

Putting this into (12.10) gives

$$
N(v)\,\mathrm{d}v \propto v^2 e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v\,, \tag{12.12}
$$

which can be normalised as before to arrive at (12.9) again.

On a side note, with $E = 1/2\, mv^2$ implying $\mathrm{d}E = mv\,\mathrm{d}v$, we can also write

$$v^2\,\mathrm{d}v = \frac{v^2\,\mathrm{d}E}{mv} = \frac{v\,\mathrm{d}E}{m} \propto \sqrt{E}\,\mathrm{d}E\,, \tag{12.13}$$

as was found in Section 3.1. This gives a spread over energies of

$$N(E)\,\mathrm{d}E \equiv N(v)\,\mathrm{d}v \propto \sqrt{E}\,e^{\frac{-E}{kT}}\,\mathrm{d}E\,. \tag{12.14}$$

12.3 Representative Speeds of Gas Particles

There are different ways of producing a representative speed of the particles. Four standard ones are derived from the Maxwell speed distribution $N(v)$ in (12.9). They are not all equally important, but are different ways of approaching the idea of a representative value. Thus it's useful to examine each briefly. In order of increasing size, they are

Most likely speed $\hat{v}$. This is found by solving $N'(\hat{v}) = 0$. The straightforward differentiation leads to

$$\hat{v} = \sqrt{\frac{2kT}{m}} = \sqrt{\frac{2RT}{M_{\mathrm{mol}}}} \simeq 1.4\,\sqrt{\frac{RT}{M_{\mathrm{mol}}}}\,, \tag{12.15}$$

where, as usual, R is the gas constant and M_{mol} is the gas particles' molar mass.

Median speed v_m. This is the speed at which half the particles are travelling slower, and half faster. Obtain it by solving

$$\int_0^{v_m} N(v)\,\mathrm{d}v = \frac{N_{\mathrm{tot}}}{2}\,. \tag{12.16}$$

Do this with a change of variables: $x \equiv \sqrt{\frac{m}{2kT}}\,v$, along with

$$\int x^2 e^{-x^2}\,\mathrm{d}x = \frac{\sqrt{\pi}}{4}\,\mathrm{erf}\,x - \frac{x}{2}e^{-x^2}\,, \tag{12.17}$$

to arrive at (with $x_m \equiv \sqrt{\frac{m}{2kT}}\,v_m$)

$$\mathrm{erf}\,x_m - \frac{2}{\sqrt{\pi}}\,x_m e^{-x_m^2} = \frac{1}{2}\,. \tag{12.18}$$

This is solved numerically, resulting in a median speed of

$$v_m \simeq \sqrt{\frac{2.366\,kT}{m}} \simeq 1.5\,\sqrt{\frac{RT}{M_{\mathrm{mol}}}}\,. \tag{12.19}$$

Mean speed $\bar{v}$ or $\langle v \rangle$.

$$\bar{v} = \int_0^\infty v\,\mathrm{prob}(v) = \int_0^\infty v\,\frac{N(v)\,\mathrm{d}v}{N_{\mathrm{tot}}} = \sqrt{\frac{2}{\pi}}\left(\frac{m}{kT}\right)^{3/2}\int_0^\infty v^3\,e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v\,. \tag{12.20}$$

This integral can be done by parts, writing it as $\int_0^\infty v^2\,v\,e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v$. We get

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}} \simeq 1.6\,\sqrt{\frac{RT}{M_{\mathrm{mol}}}}\,. \tag{12.21}$$

RMS speed $v_{\mathbf{rms}}$. This is the "(square) root (of the) mean (of the) square (of the) speed", so

$$v_{\mathrm{rms}}^2 = \langle v^2 \rangle = \int_0^\infty v^2\,\frac{N(v)\,\mathrm{d}v}{N_{\mathrm{tot}}} = \sqrt{\frac{2}{\pi}}\left(\frac{m}{kT}\right)^{3/2}\int_0^\infty v^4\,e^{\frac{-mv^2}{2kT}}\,\mathrm{d}v\,. \qquad (12.22)$$

Do this integral in the same way as (12.5)–(12.6), but now differentiate *twice* under the integral sign. The final result is (remembering to take the square root)

$$v_{\mathrm{rms}} = \sqrt{\frac{3kT}{m}} \simeq 1.7\sqrt{\frac{RT}{M_{\mathrm{mol}}}}\,. \qquad (12.23)$$

This value makes good sense, since it implies that the mean value of a particle's energy is

$$\langle E \rangle = \langle \nicefrac{1}{2}\,mv^2 \rangle = \frac{m}{2}\langle v^2 \rangle = \frac{mv_{\mathrm{rms}}^2}{2} = \frac{m}{2}\frac{3kT}{m} = \nicefrac{3}{2}\,kT\,, \qquad (12.24)$$

just as we expect from the calculation of Section 10.4: a particle's *average* thermal energy per degree of freedom for a system contacting a bath is $\nicefrac{1}{2}\,kT$, and we are only concerned with translational kinetic energy here, so there are 3 degrees of freedom. This simple connection with the average energy makes the rms speed perhaps the most widely used representative speed of the particles.

13 Theory of Transport Processes

Here we study several processes based on the concept of the mean free path, that combine to allow experiments to be done that test the validity of the models we have been using.

13.1 Mean Free Path of Gas Particles

Consider a gas with all particles alike but distinguishable, for which the Maxwell velocity distribution will apply. We wish to calculate λ, the particles' *mean free path*, or the mean distance between successive collisions. The following is a rough argument for how to do this. It can be made more rigorous by examining more closely the relevant interactions, but for our purpose it's fine. We assume the gas isn't too dense, so that each particle spends most of its time in free flight. This is a very good approximation for all manner of gases—even pseudo gases such as conduction electrons moving about in a metal, that we'll consider later.

Suppose a particle travels at mean speed $\bar{v}$ for a time Δt. It "carves out" a tube of length $\bar{v}\Delta t$. If the particles it encountered were all at rest in the laboratory (and so moving with relative speed $\bar{v}$ past our particle), the number of collisions would equal the number of particles whose centres are in this tube. For a particle density n, this number of particles (collisions) equals

$$n \times \text{volume of tube} = n\sigma\,\bar{v}\Delta t\,, \qquad (13.1)$$

where σ is the *collision cross section*. We can imagine that any particle at a distance farther than $2r$ (where r is the radius of each particle) will not be struck, so $\sigma = \pi(2r)^2$.

In this case then, the mean free path would be approximately the tube length divided by the number of collisions, or

$$\lambda \simeq \frac{\bar{v}\Delta t}{n\sigma\,\bar{v}\Delta t} = \frac{1}{n\sigma}\,. \tag{13.2}$$

In practice the particles are not at rest, so they don't come past our particle with a relative speed $\bar{v}$. Suppose instead that they all pass our particle with a relative speed $\bar{u}$, and we will disregard the finer points of averaging over the various directions from which they came. Then the number of collisions, or particles encountered, by our particle is *as if* our particle were travelling at $\bar{u}$ for a time Δt through stationary particles. In that case (and remembering that the tube length is still $\bar{v}\Delta t$),

$$\text{number of collisions} = n\sigma\,\bar{u}\Delta t\,, \quad \text{so} \;\; \lambda \simeq \frac{\bar{v}\Delta t}{n\sigma\,\bar{u}\Delta t} = \frac{\bar{v}}{n\sigma\bar{u}}\,. \tag{13.3}$$

What is $\bar{u}$? Although we won't do so here, we can use the Maxwell velocity distribution to write the combined probability that some particle ("particle 1") has velocity $\boldsymbol{v}_1$ and another particle ("2") has velocity $\boldsymbol{v}_2$. We can then write each velocity in terms of the particles' relative velocity $\boldsymbol{u} = \boldsymbol{v}_1 - \boldsymbol{v}_2$ and their centre-of-mass velocity, $\boldsymbol{V} = \frac{1}{2}\,(\boldsymbol{v}_1 + \boldsymbol{v}_2)$. We then integrate over $\boldsymbol{V}$ to get the distribution of $\boldsymbol{u}$, and find the mean of the relative speed u in the same way as was done in Section 12.3. There is a change of 6 variables $v_{1x}, \ldots, v_{2z}$ to $V_x, \ldots, u_z$ needed (requiring the determinant of a 6×6 matrix, but that turns out to be straightforward), and some more integration. We'll omit the details and simply write the answer:

$$\bar{u} = \bar{v}\sqrt{2}\,. \tag{13.4}$$

Hence (13.3) becomes

$$\lambda \simeq \frac{1}{n\sigma\sqrt{2}}\,. \tag{13.5}$$

We can also write down the *collision frequency*:

$$\text{collision frequency} \equiv \frac{\text{no. of collisions}}{\Delta t} = \frac{n\sigma\,\bar{u}\Delta t}{\Delta t} = n\sigma\bar{v}\sqrt{2}\,. \tag{13.6}$$

Example: What are the mean free path and collision frequency for particles in the air of our lecture theatre?

Give these particles a temperature of $300\,\text{K}$ and a pressure of $10^5\,\text{Pa}$. Use the ideal gas law $PV = NkT$ to write the particle density as

$$n = \frac{\text{no. of molecules in } V}{V} = \frac{N}{V} = \frac{P}{kT} \tag{13.7}$$

and

$$\sigma = 4\pi r^2\,, \quad \text{where } r \simeq 10^{-10}\,\text{m}. \tag{13.8}$$

So the mean free path is

$$\lambda = \frac{1}{n\sigma\sqrt{2}} = \frac{kT}{P4\pi r^2\sqrt{2}} = \frac{1.38\underline{\overline{-23}} \times 300}{10^5 \times 4\pi \times 10^{-20} \times \sqrt{2}}\,\text{m} \simeq 0.23\;\mu\text{m}. \qquad \underline{\underline{\text{Answer}}} \tag{13.9}$$

The collision frequency is

$$n\sigma\bar{v}\sqrt{2} = \frac{P}{kT}\,4\pi r^2\,\sqrt{\frac{8kT}{\pi m}}\,\sqrt{2} = \frac{P}{kT}\,4\pi r^2\,\sqrt{\frac{16RT}{\pi M_{\text{mol}}}}\,. \tag{13.10}$$

For air, $M_{\text{mol}} \simeq 0.8 \times 28 + 0.2 \times 32 = 28.8\,\text{g}$, so the collision frequency is

$$\frac{10^5}{1.38 \underline{\underline{-23}} \times 300} \times 4\pi \times 10^{-20} \sqrt{\frac{16 \times 8.314 \times 300}{\pi \times 0.0288}} \simeq 2.0 \times 10^9 \text{ s}^{-1}. \tag{13.11}$$

Two thousand million collisions per second is an extraordinary number, but there are, of course, many air molecules in the room.

13.2   Viscosity

Our discussion of mean free path in a gas can be applied to connect ideas of viscosity, thermal conductivity, and heat capacity for a gas. More generally it can also be applied to a fluid (being a substance that takes the shape of its container), provided the fluid's internal processes are not so complex as to negate the assumptions that we'll make.

Consider a fluid with, say, a metal plate on top in the xy plane that is pulled with some constant velocity. The plate will drag fluid particles along with it. For gases, a good model of the situation is that the gas is composed of a stack of plates in the xy plane, where the plate at $z = $ constant experiences a force that drags it against internal friction along the x direction with velocity $u_x(z)$. As we pull on the top plate, excess x momentum is transferred by random particle motion down through the plates, which drags *them* in turn—although the lower the plates are, the lesser the x velocity they'll inherit. (We disregard the ever-present mean x momentum, which must cancel over large areas, otherwise the plates would move spontaneously. The only x momentum used below is the excess responsible for the plates moving.)

We wish to relate the force required to drag the top plate to some measure of the gas's viscosity. First, note that

$$\text{force applied} = \frac{\text{momentum transferred}}{\text{time taken}}. \tag{13.12}$$

So as the plate at $z = $ constant is dragged in the x direction, we can define a quantity sometimes called T^{xz} by

$$T^{xz} \equiv \begin{bmatrix} x \text{ component of force needed} \\ \text{to drag the unit-area plate of} \\ \text{gas at } z \text{ with constant velocity} \end{bmatrix} \overset{(13.12)}{=\!=\!=} \frac{p_x \text{ transferred to plate}}{\text{area} \times \text{time taken}}$$

$$= \text{net } p_x \text{ transferred up through plate at } z, \text{ per unit area per unit time}$$

$$= (p_x \text{ per particle}) \times \begin{bmatrix} \text{no. of particles transferred up} \\ \text{through plate at } z, \text{ per unit area} \\ \text{per unit time} \end{bmatrix}. \tag{13.13}$$

With n particles per unit volume, consider that $n/3$ have some motion in the z direction, with half of these going up and half going down. So the $n/6$ going down carry a higher x momentum down (which accelerates those bottom layers), and the $n/6$ going up carry a lower x momentum up. Thus x momentum leaks down through the plates, pulling them to the right with decreasing force the lower it goes.

The number of particles passing, say, up through the z plate per unit area per unit time is their *flux density* in the z direction. (Flux density is very often just called flux, but

this conflicts with its most well-known use, in electromagnetism. So we won't call it flux.)
If we consider N particles per unit volume passing with speed v for a time Δt along a tube
of cross-sectional area A, then the flux density through the tube's end face is governed by
how many particles passed through the face in this time:

$$\text{flux density} = \frac{\text{no. through face}}{\text{area} \times \text{time}} = \frac{NAv\Delta t}{A\Delta t} = Nv \,. \tag{13.14}$$

As used here, "flux" is synonymous with *current*. Flux can refer to the motion of anything, such as
particles, mass, electric charge, as well as being applied to field lines in electromagnetism. More
generally, flux density is a vector that can be dotted with the normal to an area to tell us the flux
through that area. For the flow of a substance, the expression

$$\boxed{\text{flux density} = \text{substance density} \times \text{substance velocity}} \tag{13.15}$$

is well worth remembering. Note that the first word "density" in (13.15) refers to the unit spatial
area, not the unit time; that is, flux density is flux per unit area, where flux means how much
substance flows per unit time. So flux equals flux density times an area (not times a time). The
second word "density" in (13.15) refers to the amount of substance per unit volume.

In our case, $N = n/6$ in each direction across a plate, and the flux density in each
direction is then $n\bar{v}/6$. So (13.13) says that the average "up" contribution to T^{xz} is
$n\bar{v}/6 \times (p_x$ per particle). The p_x carried by each particle is what that particle inherited
at its last collision, which happened a distance of approximately λ from the plate at z.
For "up" motion (say, the direction of increasing z), this momentum is formed from the
particle's velocity at $z - \lambda$, so is approximately $mu_x(z - \lambda)$. Similarly, the momentum
carried down is $mu_x(z + \lambda)$. The net momentum travelling up through the plate is then

$$T^{xz} = \underbrace{\frac{n\bar{v}}{6}mu_x(z - \lambda)}_{\text{"up" part}} - \underbrace{\frac{n\bar{v}}{6}mu_x(z + \lambda)}_{\text{"down" part}}$$

$$\simeq \frac{n\bar{v}m}{6}\left[\cancel{u_x(z)} - u'_x(z)\lambda - \cancel{u_x(z)} - u'_x(z)\lambda\right] \quad \text{(a Taylor expansion)}$$

$$= \frac{-n\bar{v}m\lambda}{3}\frac{\partial u_x}{\partial z} \equiv -\eta\frac{\partial u_x}{\partial z} \,, \tag{13.16}$$

where we have written a partial derivative to show that more generally u_x depends on y
as well, and where $\eta \equiv n\bar{v}m\lambda/3$ is the *coefficient of viscosity*. Since our top metal plate is
only moving in the x direction, T^{xz} is the whole force per unit plate area needed to drag
it, and so η can be measured.

Writing η in terms of what we know already,

$$\eta = \frac{n}{3}\bar{v}m\lambda = \frac{n}{3}\sqrt{\frac{8RT}{\pi M_{\text{mol}}}}\frac{M_{\text{mol}}}{N_A}\frac{1}{n\sigma\sqrt{2}} = \frac{\sqrt{4RTM_{\text{mol}}}}{3\sqrt{\pi}N_A\,\sigma} \,. \tag{13.17}$$

Note that the gas's particle density n cancels; the surprising result is that viscosity is
independent of this density (at a given temperature). This was both derived and experi-
mentally confirmed by Maxwell.

A gas of close-packed particles each with radius r has a mass density in the region of

$$\frac{m}{r^3} = \frac{M_{\text{mol}}}{N_A\,r^3} \,. \tag{13.18}$$

Equations (13.17) and (13.18) provided the first values of r and N_A by Loschmidt in 1885.

Equation (13.17) predicts, correctly, that the viscosity of a gas increases with temperature. The viscosity of *liquids* actually decreases with temperature. For liquids, we must add something to the model: the particles are so close together that particles on adjacent planes will interact with each other, even though they are not crossing the planes. But that's outside this course.

13.3   Thermal Conductivity

In Section 7 we looked at thermal conduction from a bulk point of view. Now we investigate the coefficient of thermal conductivity κ with an atomic approach almost identical to the above for viscosity. In particular, consider a plate in the xy plane at some temperature T. The flux density of thermal energy across a plate at $z = $ constant is the z component of (7.1), or

$$J_z = -\kappa \frac{\partial T}{\partial z} \,. \tag{13.19}$$

In fact J_z has another name: T^{tz}, where T^{tz} and T^{xz} are 2 of 10 components of the *stress–energy tensor* (which is not part of our course). Suffice it to say that if we arrange for the speed of light to be dimensionless, then the components of the stress–energy tensor all have units of pressure. Einstein postulated that stress–energy curves spacetime; that is, stress–energy is to Einstein what mass is to Newton: the source of gravity.

Equation (13.19) holds well for liquids and gases, as well as solids with sufficient homogeneity. It's analogous to (13.16); for viscosity we looked at the transfer of x momentum, but now we look at the transfer of thermal energy E. Particles at height z each have thermal energy $E(z)$. We have

$$J_z = \text{net heat transferred up through plate at } z, \text{ per unit area per unit time}$$

$$= (\text{heat energy per particle}) \times \left[\begin{array}{l} \text{no. of particles transferred up} \\ \text{through plate at } z, \text{ per unit area} \\ \text{per unit time} \end{array} \right]$$

$$= \underbrace{\frac{n\bar{v}}{6} E(z - \lambda)}_{\text{``up''}} - \underbrace{\frac{n\bar{v}}{6} E(z + \lambda)}_{\text{``down''}}$$

$$\simeq \frac{n\bar{v}}{6} \left[\cancel{E(z)} - E'(z)\lambda - \cancel{E(z)} - E'(z)\lambda \right]$$

$$= \frac{-n\bar{v}\lambda}{3} \frac{\partial E}{\partial z} \,, \tag{13.20}$$

where again we have used a partial derivative for generality. But if we just consider heat conduction in the z direction, we can use an ordinary derivative, and equate (13.19) with (13.20):

$$-\kappa \frac{\mathrm{d}T}{\mathrm{d}z} = \frac{-n\bar{v}\lambda}{3} \frac{\mathrm{d}E}{\mathrm{d}z} \,. \tag{13.21}$$

But that means

$$\kappa = \frac{n\bar{v}\lambda}{3} \frac{\mathrm{d}E}{\mathrm{d}T} = \frac{n\bar{v}\lambda}{3} C_V^{1\,\text{particle}} = \frac{n\bar{v}\lambda}{3} \frac{C_V^{\text{mol}}}{N_A} \,. \tag{13.22}$$

Now use this last equation along with $\eta = n\bar{v}m\lambda/3$ to write

$$\frac{\kappa}{\eta} = \frac{C_V^{\text{mol}}}{M_{\text{mol}}} = C_V^{\text{sp}}. \tag{13.23}$$

So we have related viscosity, thermal conductivity, and heat capacity using an atomic view of matter. Experiments yield values of

$$\frac{\kappa}{\eta} \approx (1.5 \to 2.5) \times C_V^{\text{sp}}. \tag{13.24}$$

Their approximate agreement with theory forms a good justification for the validity of the kinetic/atomic models that we have been using. We can easily expect to be out by a factor of 2 in our calculations, as these are based on heuristic models with a heavy reliance on averaging. But we haven't done too badly.

14 Bands, Levels, and States

Discrete values of energy that a system can have are called *energy levels*, and each one may contain some degeneracy; that is, there might be several states all with the same energy. When energy levels are very close together so as to form more or less a continuum, the language used to describe them can be a little confusing. Such a set of such energy levels is called an energy *band*, and the levels and states comprising it are treated together and collectively just called *energy states*.

As an example, consider two energy levels. Suppose that for a system in contact with a bath at 17°C ($kT = 0.025$ eV), level 1 has energy $E_1 = 1$ eV and degeneracy $g_1 = n$ states. Level 2 has $E_2 = 1.02$ eV and $g_2 = n$ states also. What is the ratio of the numbers of particles in these two levels? If there are N_{tot} particles present, then the number of particles in level i is

$$N_i = N_{\text{tot}} \times \text{probability for each particle to have energy } E_i$$
$$= N_{\text{tot}} \times \sum_{\text{states}} \text{prob. for a particle to be in a state at level } E_i$$
$$= N_{\text{tot}}\, g_i\, C \exp \frac{-E_i}{kT}, \tag{14.1}$$

where C is a normalisation. Then

$$\frac{N_2}{N_1} = \frac{N_{\text{tot}}\, n\, C \exp \frac{-1.02}{0.025}}{N_{\text{tot}}\, n\, C \exp \frac{-1.00}{0.025}} = \exp \frac{-0.02}{0.025} \simeq 0.45. \tag{14.2}$$

Suppose now that the n states in level 2 are spread out, forming an energy band from $1.02 \to 1.03$ eV. (These energy states have now become energy levels!) Now what is the ratio of numbers of particles in band 2 to level 1 (a.k.a. band 1)? Let's first approximate band 2 as *two* energy levels (at 1.02 and 1.03 eV), each with degeneracy $n/2$ states. Thus

$$g_1 = n, \quad g_{1.02} = g_{1.03} = n/2. \tag{14.3}$$

Then

$$\frac{N_2}{N_1} = \frac{N_\text{tot} \frac{n}{2} \, C \exp \frac{-1.02}{0.025} + N_\text{tot} \frac{n}{2} \, C \exp \frac{-1.03}{0.025}}{N_\text{tot} \, n \, C \exp \frac{-1.00}{0.025}}$$

$$= \frac{1}{2} \exp \frac{-0.02}{0.025} + \frac{1}{2} \exp \frac{-0.03}{0.025} \simeq 0.375 \,. \tag{14.4}$$

Now let's do better by approximating band 2 as *three* energy levels (at 1.02, 1.025, 1.03 eV), each with $n/3$ states:

$$g_1 = n \,, \quad g_{1.02} = g_{1.025} = g_{1.03} = n/3 \,. \tag{14.5}$$

Then

$$\frac{N_2}{N_1} = \frac{N_\text{tot} \frac{n}{3} \, C \exp \frac{-1.02}{0.025} + N_\text{tot} \frac{n}{3} \, C \exp \frac{-1.025}{0.025} + N_\text{tot} \frac{n}{3} \, C \exp \frac{-1.03}{0.025}}{N_\text{tot} \, n \, C \exp \frac{-1.00}{0.025}}$$

$$= \frac{1}{3} \exp \frac{-0.02}{0.025} + \frac{1}{3} \exp \frac{-0.025}{0.025} + \frac{1}{3} \exp \frac{-0.03}{0.025} \simeq 0.373 \,. \tag{14.6}$$

Suppose we could spread the n energy levels out evenly over the band. Then

$$N_2 = N_\text{tot} \times \text{prob. for a particle to be in band 2}$$

$$= N_\text{tot} \sum_{\text{levels}} \text{prob. for a particle to be in a level } E$$

$$= N_\text{tot} \sum_{\text{levels}} (\text{prob. for a particle to be in a state around level } E)$$

$$\times (\text{number of states around level } E) \,. \tag{14.7}$$

Imagine that in fact we spread the n levels out over the band's 0.01 eV width to form a *continuum*. Then the number of states (or levels!) around E is

$$g(E) \, \mathrm{d}E = \frac{n}{0.01 \text{ eV}} \, \mathrm{d}E \,. \tag{14.8}$$

For such a case, (14.7) becomes

$$N_2 = N_\text{tot} \int C e^{\frac{-E}{kT}} \, g(E) \, \mathrm{d}E$$

$$= N_\text{tot} \frac{n}{0.01 \text{ eV}} \, C \int_{1.02 \text{ eV}}^{1.03 \text{ eV}} e^{\frac{-E}{kT}} \, \mathrm{d}E$$

$$= N_\text{tot} \frac{n}{0.01 \text{ eV}} \, C(-0.025 \text{ eV}) \left[\exp \frac{-1.03}{0.025} - \exp \frac{-1.02}{0.025} \right] \,, \tag{14.9}$$

and with N_1 unchanged from (14.2), we get

$$\frac{N_2}{N_1} = \frac{-0.025}{0.01} \left[\exp \frac{-1.03}{0.025} - \exp \frac{-1.02}{0.025} \right] \simeq 0.370 \,. \tag{14.10}$$

More generally, (14.7) can be written as

$$\text{number of particles in } E_0 \to E_0 + \Delta E \; = \; N_{\text{tot}} \int_{E_0}^{E_0+\Delta E} \underbrace{C e^{\frac{-E}{kT}}}_{\substack{\text{prob. to be in} \\ \text{state around } E}} \; \times \; \underbrace{g(E)\,\mathrm{d}E}_{\substack{\text{number of states} \\ \text{around } E}} \tag{14.11}$$

or

$$\underbrace{\text{number of particles in } E \to E + \mathrm{d}E}_{\substack{\equiv N(E)\,\mathrm{d}E \\ \text{number of particles per} \\ \text{unit energy interval}}} \; = \; \underbrace{N_{\text{tot}}\, C\, e^{\frac{-E}{kT}}}_{\substack{= \bar{n}(E),\text{ mean} \\ \text{number of} \\ \text{particles/state}}} \; \times \; \underbrace{g(E)\,\mathrm{d}E}_{\substack{\text{number of states} \\ \text{per unit energy} \\ \text{interval}}} \tag{14.12}$$

This is written more compactly as

$$N(E) = \bar{n}(E)\, g(E)\,. \tag{14.13}$$

Systems are characterised by $\bar{n}(E)$, the mean number of particles per state (or *occupation number*), and $g(E)$, the density of states, or "spectrum of accessible states". When the number of particles is much less than the number of states, the Boltzmann distribution holds, meaning

$$\bar{n}(E) \propto e^{\frac{-E}{kT}}\,. \tag{14.14}$$

In contrast to the occupation number, the density of states $g(E)$ varies widely from system to system.

15 Introducing Quantum Statistics

In quantum mechanics, identical particles really are identical: they cannot be distinguished, even in principle. This loss of individuality has consequences for arguments in which they are counted. Previously we have focused on a single particle and asked for the chance that it can occupy any one of several different states (and so calculated the mean number of particles per state). We must now take two things into account:

- the particles' indistinguishability,

- the higher particle densities encountered in systems for which a quantum mechanical treatment is necessary.

These are both incorporated by shifting focus to a *state*, and computing the chance that it's occupied by some number of particles. In (10.19) we wrote down the probability that a single quantum state of energy E_s and volume V_s is occupied by N_s particles:

$$p \propto \exp \frac{-E_s - PV_s + \mu\, N_s}{kT}\,. \tag{15.1}$$

Typically the volume of a state is fixed; the state is occupied by n particles, each of energy E. Then (15.1) becomes

$$p_n \propto \exp \frac{-nE + \mu\, n}{kT} = \exp \frac{-n(E - \mu)}{kT}\,. \tag{15.2}$$

Write this last equation as

$$p_n = \frac{e^{-n\alpha}}{Z}, \quad \text{where } \alpha \equiv \frac{E - \mu}{kT} > 0, \tag{15.3}$$

and Z is the (reciprocal of the) normalisation, called the *partition function*, that we met previously in Section 11.2. Determine it by writing

$$1 = \sum_n p_n = \frac{\sum_n e^{-n\alpha}}{Z}, \tag{15.4}$$

so that

$$Z = \sum_n e^{-n\alpha}. \tag{15.5}$$

The occupation number, being the mean number of particles in a state, is

$$\bar{n} = \sum_n n\, p_n = \frac{\sum_n n e^{-n\alpha}}{Z} = \frac{-1}{Z}\frac{\partial Z}{\partial \alpha}. \tag{15.6}$$

This is an example of the utility of the partition function: once it's found, other quantities can be produced from it by simple operations such as differentiation.

15.1   Two Types of Fundamental Particle

Experiments show that fundamental particles come in either of two types:

Fermions. At most one fermion can occupy a given state, so (15.5) becomes

$$Z = \sum_{n=0}^{1} e^{-n\alpha} = 1 + e^{-\alpha}, \tag{15.7}$$

Hence (15.6) gives

$$\bar{n} = \frac{e^{-\alpha}}{1 + e^{-\alpha}} = \frac{1}{e^{\alpha} + 1}. \tag{15.8}$$

Bosons. Any number of bosons can occupy a given state, so (15.5) becomes

$$Z = \sum_{n=0}^{\infty} e^{-n\alpha} = \frac{1}{1 - e^{-\alpha}}, \quad \text{and } \bar{n} = \frac{1}{e^{\alpha} - 1}. \tag{15.9}$$

The occupation number can be written for both particles at once with

$$\boxed{\bar{n} = \frac{1}{\exp\frac{E-\mu}{kT} \pm 1} \qquad \begin{cases} \text{fermions} \\ \text{bosons} \end{cases}} \tag{15.10}$$

Fermions turn out be particles with odd half-integral spin, such as electrons, positrons, protons, neutrons, neutrinos, and muons. Bosons turn out to be particles with integer spin, such as α particles, pions, photons, and deuterons. Fundamental particles don't

seem to exist with any other spin varieties than these two, so they are all either fermions or bosons. Just why spin should determine a particle's occupation number is analysed by relativistic quantum mechanics, but the fundamental reason is not well understood. The study of fermions is known as *Fermi–Dirac statistics*, while that of bosons is *Bose–Einstein statistics*. In the high-energy limit, or low particle-density limit, these both reduce to what we have been studying up until now, *Maxwell–Boltzmann statistics*.

16   Blackbody Radiation

Hot objects contain oscillating charges, and oscillating charges radiate electromagnetic waves. An obvious example is the light emitted by the hot gas that makes up a flame. The electromagnetic theory of just how this occurs is complicated even for very simple systems; but given that a huge number of oscillating charges are responsible for the light emitted, it turns out that we can use statistical mechanics to examine such macroscopic objects within electromagnetic theory.

But the emission of light is not an equilibrium process; there's a continuous transfer of energy from the object to the waves, which is then lost from the object. We have only considered equilibrium processes in this course, and the subject of non-equilibrium processes is an advanced branch of statistical mechanics. However, we can calculate how much light is emitted from a hot body by considering a related process that *does* occur in equilibrium. The idea that allows us to make this connection is the *Principle of Detailed Balance*, which is discussed in [3]. Because oscillating charges emit light, if a body has charges that resonate at some particular frequency, then not only will it readily *emit* light of that frequency, but it will also readily *absorb* light of that frequency. The Principle of Detailed Balance postulates that the ability to emit equals the ability to absorb:

> When in thermal equilibrium with a bath of electromagnetic waves, any object—regardless of its colour or makeup—emits the same spectrum and intensity that it absorbs.

So we will derive the spectrum emitted by a hot body by examining a related scenario: the spectrum that exists inside a hot oven. The walls inside the oven are in equilibrium with the radiation inside, so we can use statistical mechanics to examine *that*, and transfer what we learn to the emitting hot body.

Consider, then, a perfectly absorbing ("black") body placed in an oven that is ideal in the sense that it is "perfectly emitting" (which we'll define in a moment). The black body must emit exactly what it absorbs; but by definition, it absorbs all the radiation it receives. So it must emit the same spectrum that the oven produces. But the mechanism for how the body emits doesn't depend on the oven, so the black body must therefore also emit identically when *outside* the oven. We conclude that the spectrum of frequencies produced by a black body equals that found inside an ideal oven.

16.1   The Radiation Inside an Oven

Next consider an oven (often called a cavity) with hot walls, containing radiation. We wish to find this radiation's *spectral energy density* $\varrho(f)$: its total electromagnetic energy

per unit frequency per unit oven volume. Different hot materials emit different amounts of each wavelength, so we cannot hope to use only general arguments to obtain the spectral energy density of an oven made from some arbitrary material. Also, we cannot expect to discuss the emission of arbitrarily low frequencies (long wavelengths) from any one oven, it being problematic to discuss a light wave of longer wavelength than the typical size of the object that produced it.

What electromagnetic frequencies exist inside the oven? It might be argued that the electric field inside a metal oven will go to zero at the walls since otherwise wall currents would arise which would then eliminate the field at the walls. That would quantise the field modes so that only certain frequencies would be present—although for all practical purposes they would form a continuum. But in a ceramic oven, the field need not go to zero at the walls, and so the frequencies present need not be quantised. On the other hand, if the walls inside an oven are reflective enough that a light wave inside bounces back and forth many times, it will be reinforced if a whole number of wavelengths fit into a round trip. Different ovens will have different amounts of internal reflectivity, and different-sized ovens will reinforce some wavelengths but not others.

The task of calculating a spectral energy density is beginning to look difficult! To make progress, we consider an idealised oven that holds a continuum of wavelengths. Its wall oscillators produce light of all frequencies; this light bounces about inside the oven, sometimes reflected and sometimes not, so that the spread of frequencies quickly tends toward some equilibrium distribution.

The following argument suggests that the oven's shape can be arbitrary. Join two differently shaped idealised ovens at the same temperature via a hole. If the radiation spectra of the two differed around some particular frequency (say, yellow light), we could presumably introduce a filter that passed that frequency only. That would allow a flow of energy in one direction through the hole, which would perhaps act to "unequalise" the temperatures. It's unreasonable for the system to depart from thermal equilibrium in such a way—it contravenes the Second Law of Thermodynamics. So we might conclude that there cannot be such a flow of energy, so that the oven's shape doesn't affect the spectrum of radiation inside.

Actually, this argument is not quite as straightforward as it might appear. While the filter would pass yellow light into the oven whose walls did not naturally emit much yellow light, the Principle of Detailed Balance says that those walls wouldn't absorb much yellow light either, in which case that oven's temperature need not increase. Would the yellow light then build up inside that oven, perhaps interacting with the filter to heat it up until it broke down? Also, the hot filter would emit radiation of its own. We will simply postulate that an idealised oven's spectral energy density is independent of its shape, and appeal to experiment for validation.

In that case, consider an oven shaped as a rectangular box with side lengths L_x, L_y, L_z and volume $V = L_x L_y L_z$. We assume that

– the oven walls are continuously emitting and absorbing radiation,

– the oven's shape doesn't affect its spectrum,

– there is no restriction on what frequencies can exist inside the oven,

– the walls contain a huge number of quantised harmonic oscillators,

– at thermal equilibrium, the total energy of the oven radiation in one "frequency state" (defined soon) in the frequency interval $f \to f + \mathrm{d}f$ equals the mean thermal energy $\varepsilon(f)$ of a wall oscillator at frequency f. That is,

$$\begin{bmatrix}\text{total energy of radiation in } f \to f + \mathrm{d}f\end{bmatrix} \longmapsto \quad \frac{\varrho(f)\,\mathrm{d}f\,V}{g(f)\,\mathrm{d}f} = \varepsilon(f)\,, \qquad (16.1)$$

in which case

$$\varrho(f) = \frac{\varepsilon(f)\,g(f)}{V}\,. \qquad (16.2)$$

We calculate the mean thermal energy $\varepsilon(f)$ and the density of states $g(f)$ as follows.

First Requirement: the Mean Thermal Energy of an Oscillator, $\varepsilon(f)$

Model the walls of the oven as composed of a set of quantum oscillators held at temperature T. (That is, we can consider the walls to be in contact with a heat bath at T.) Recall that the nth energy level of the quantised oscillator has energy $(n + 1/2)\,hf$, giving it a *thermal* energy of nhf, since the $1/2\,hf$ is the ever-present zero-point energy which cannot be taken away from the oscillator; it's not thermal energy, so does not enter our analysis. Refer to Section 10.3.2 to write the mean thermal energy of the oscillators as

$$\varepsilon(f) = \frac{\sum_{n=0}^{\infty} e^{\frac{-nhf}{kT}} nhf}{\sum_{n=0}^{\infty} e^{\frac{-nhf}{kT}}} = \frac{hf \sum e^{n\alpha} n}{\sum e^{n\alpha}}\,, \quad \text{with } \alpha \equiv \frac{-hf}{kT}\,. \qquad (16.3)$$

Because $e^{\alpha} < 1$, the middle denominator in (16.3) is simply a geometric series:

$$\sum_{n=0}^{\infty} e^{n\alpha} = \frac{1}{1 - e^{\alpha}}\,, \qquad (16.4)$$

in which case the sum in the middle numerator in (16.3) is

$$\sum_{n=0}^{\infty} e^{n\alpha} n = \frac{\mathrm{d}}{\mathrm{d}\alpha} \sum_{n=0}^{\infty} e^{n\alpha} = \frac{\mathrm{d}}{\mathrm{d}\alpha} \frac{1}{1 - e^{\alpha}} = \frac{e^{\alpha}}{\left(1 - e^{\alpha}\right)^2}\,. \qquad (16.5)$$

Hence (16.3) becomes

$$\varepsilon(f) = \frac{hf}{e^{\frac{hf}{kT}} - 1}\,. \qquad (16.6)$$

Note that in the regimes of low and high temperature,

$$kT \ll hf \Longleftrightarrow \varepsilon(f) \to 0$$
$$kT \gg hf \Longleftrightarrow \varepsilon(f) \simeq kT\,. \qquad (16.7)$$

The first equation just shows that the thermal energy vanishes as $T \to 0$, and the second is the expected result for a classical oscillator (2 d.o.f.), from the Equipartition theorem.

The Schrödinger equation ascribes an infinite number of energy levels to a harmonic oscillator: the nth level has thermal energy nhf. The quantum statistics view of Section 15 is that the oscillator has just one state, which can be occupied by any number n of particles,

called *photons*, that each have energy hf. The expression for $\varepsilon(f)$ in (16.6) implies that the mean number of photons of energy hf in the oven is

$$\bar{n}(f) = \frac{\text{mean energy of photons}}{\text{energy per photon}} = \frac{\varepsilon(f)}{hf} = \frac{1}{e^{\frac{hf}{kT}} - 1} . \qquad (16.8)$$

Compare this with (15.10): we conclude that photons are bosons with chemical potential $\mu = 0$.

Second Requirement: Density of States $g(f)$ for Light Waves

In Section 3.1 we derived an ideal gas's density of states $g(E)$ at energy E by calculating the total number of energy states in the energy range $0 \to E$, and then using $g(E) = \Omega'_{\text{tot}}(E)$. We calculated the total number of states by postulating that each state could be represented by a cell in phase space. The total number of states was then just the total number of cells, which equalled the total phase space volume able to be occupied for energies from 0 to E divided by the volume of one cell. Constructing a cell wasn't quite a unique affair, because although we used Planck's constant h to give the cell a natural size, the remarks on page 20 showed that a state itself was not very well defined. And yet, happily, this latitude in how we defined a state had no effect on calculations of entropy increase.

We'll use a similar argument here to count the number of wave states in the oven. Just as for the ideal gas above, there is some latitude in how we can define a wave state. The following line of argument is accepted in statistical mechanics because its prediction of how much radiation exists inside an oven has stood up well to experimental tests.

So, as in Section 3.1, calculate the density of states $g(f)$ at frequency f by defining and counting the total number of states in the frequency range $0 \to f$, and then use $g(f) = \Omega'_{\text{tot}}(f)$.

There is an immediate problem in defining and counting wave states. With the range of allowed frequencies being continuous, the idea of a frequency state doesn't immediately make any sense. We can make progress by identifying each wave by its "wavenumber" vector $\boldsymbol{k}$ (usually just called its wave vector), whose length is $k = 2\pi/\lambda = 2\pi f/c$ where c is the speed of light, and which points in the wave's direction of travel.

> Why use $\boldsymbol{k}$ to calculate the density of states? If $\boldsymbol{n}$ is a unit vector pointing in the wave's direction of travel, could we define a "frequency vector" $\boldsymbol{f} \equiv f\boldsymbol{n}$, or a "wavelength vector" $\boldsymbol{\lambda} \equiv \lambda\boldsymbol{n}$, and use one of these instead to characterise the wave? It turns out that these are not reasonable quantities to define. Suppose the wave's direction of travel has angle θ with the x axis. If $\boldsymbol{f}$ were indeed a vector, we would probably expect that its x component $f_x = f\cos\theta$ would be the frequency of the wave crests' intersections with the x axis; but that frequency is in fact f! Also, if $\boldsymbol{\lambda}$ were indeed a vector, we might expect that its x component $\lambda_x = \lambda\cos\theta$ would be the wavelength of the wave crests' intersections with the x axis; but this wavelength is actually $\lambda/\cos\theta$. So neither of these would-be "vectors" $\boldsymbol{f}$ or $\boldsymbol{\lambda}$ is particularly meaningful, and that's why physicists don't define them.
>
> But the appearance of $\cos\theta$ in the denominator a couple of lines up suggests that the *reciprocal* of wavelength might make a vector. And indeed it does: enter the wave vector $\boldsymbol{k} \equiv k\boldsymbol{n}$, where $k \equiv 2\pi/\lambda$ is the wave number. The 2π is just for convenience, but the λ in the denominator now sends the $\cos\theta$ back to the numerator in the previous paragraph's discussion of components. That means the wave vector's x component $k_x = k\cos\theta$ is indeed the wave number of the crests' intersections with the x axis: $k_x = 2\pi/(\text{wavelength of crests' intersections})$. So $\boldsymbol{k}$ behaves just as we expect vectors to behave. That's why it is so useful for characterising waves, and why it's found everywhere in wave theory.

Any particular wave has a vector $\boldsymbol{k} = (k_x, k_y, k_z)$. Perhaps we can find the number $\Omega_{\text{tot}}(f)$ of states in the frequency range $0 \to f$ by counting how many wave vectors can fit into a sphere in 3-dimensional "$\boldsymbol{k}$ space" such that the longest vector has a length corresponding to frequency f: this length will be $k = 2\pi/\lambda = 2\pi f/c$. But, of course, there are an infinite number of such vectors, so it will do no good to treat each one as a separate state. We must postulate something new: that the vectors can be "binned", grouped into cells in $\boldsymbol{k}$ space. Each cell defines 2 states, corresponding to the 2 possible polarisations that a wave can have. A cell is defined by requiring the coordinates (k_x, k_y, k_z) of one of its corners to satisfy a certain condition. We must search for a condition that leads to experimentally verifiable predictions. Consider two such conditions, which both lead to the same $g(f)$ (which is eventually given experimental validation).

(a) We treat the oven as holding a continuum of travelling waves, and require the coordinates k_x, k_y, k_z of a cell each to be related to a whole number of wavelengths that fit into the corresponding side lengths of the oven. That is, cells that are adjacent along say the x axis describe waves whose numbers of wavelengths fitting into a side length along the x axis differ by 1. In this case, remembering the factor of 2 for the polarisations,

$$\Omega_{\text{tot}} = 2 \times \text{number of cells}$$
$$= 2 \times \frac{\text{volume of sphere of radius } k = 2\pi/\lambda}{\text{volume of one cell}} \,. \tag{16.9}$$

What is the volume of a cell? We need to know each of its edge lengths. This is a question that arises frequently in statistical mechanics. We can answer it with the help of some apparently unrelated analysis: if we have some function $z(x, y)$

$$z = ax + by + c \tag{16.10}$$

where a, b, c are constants, how does z increase when x and y increase? By definition, increases Δx and Δy in x and y cause an increase Δz in z, so

$$z + \Delta z = a(x + \Delta x) + b(y + \Delta y) + c \,. \tag{16.11}$$

Subtracting (16.10) from (16.11) gives

$$\Delta z = a\Delta x + b\Delta y \,. \tag{16.12}$$

This is a useful expression because it means that the operation of finding the increase is *linear*. Linearity is a central theme in physics. An operation L is linear if, for constants a and b,

$$\boxed{L(ax + by) = aL(x) + bL(y) \,.} \tag{16.13}$$

It suffices to have just two terms on the right hand side of (16.13), but it's easy to show that if L is linear, then it can be applied to any number of terms:

$$L(ax + by + cz + \dots) = aL(x) + bL(y) + cL(z) + \dots \tag{16.14}$$

We'll use the idea that Δ is linear in a moment.

With Δk_x being the increase in k_x along the side of a cell in the x direction, and similarly for y and z, a cell must have volume $\Delta k_x \, \Delta k_y \, \Delta k_z$. We can calculate e.g. Δk_x by writing k_x in terms of something else with a known increase along a cell's x direction. We haven't yet used the requirement that a whole number of

wavelengths must fit into the corresponding side lengths of the oven. These whole numbers of wavelengths are triplets (n_x, n_y, n_z) such that

$$n_\alpha \lambda_\alpha = L_\alpha \,, \quad \text{for } \alpha = x, y, z \,. \tag{16.15}$$

Thus

$$k_\alpha = \frac{2\pi}{\lambda_\alpha} = \frac{2\pi n_\alpha}{L_\alpha} \,. \tag{16.16}$$

Now remembering that Δ is linear, we can immediately write

$$\Delta k_\alpha = \frac{2\pi \Delta n_\alpha}{L_\alpha} = \frac{2\pi}{L_\alpha} \,, \tag{16.17}$$

since by definition $\Delta n_x = 1$ when we move along a cell in the x direction, and similarly for y, z. So

$$\text{cell volume} = \Delta k_x \, \Delta k_y \, \Delta k_z = \frac{8\pi^3}{L_x L_y L_z} = \frac{8\pi^3}{V} \,. \tag{16.18}$$

Hence (16.9) becomes

$$\Omega_{\text{tot}} = 2 \times \frac{\frac{4}{3}\pi \left(\frac{2\pi f}{c}\right)^3}{8\pi^3/V} = \frac{8\pi f^3 V}{3c^3} \,. \tag{16.19}$$

(b) Alternatively, we treat the oven as containing standing waves only. In that case the k_x, k_y, k_z each specify that a whole number of *half*-wavelengths fits into each dimension of the oven. For this case we use only positive values of the wave vector components; the reason is because while a standing wave is comprised of superposed travelling waves of opposite-sign wavenumbers, it needs only the positive wavenumber to quantify it. Again with a factor of 2 for the polarisations,

$$\Omega_{\text{tot}} = 2 \times \text{number of cells}$$
$$= 2 \times \frac{\text{volume of one } octant \text{ of sphere of radius } k = 2\pi/\lambda}{\text{volume of one cell}} \,, \tag{16.20}$$

since positive wavenumbers comprise just one octant. Now to determine the cell volume $\Delta k_x \, \Delta k_y \, \Delta k_z$, realise that the whole numbers of wavelengths (n_x, n_y, n_z) satisfy

$$n_\alpha \frac{\lambda_\alpha}{2} = L_\alpha \,, \quad \text{for } \alpha = x, y, z \,; \tag{16.21}$$

thus

$$k_\alpha = \frac{2\pi}{\lambda_\alpha} = \frac{\pi n_\alpha}{L_\alpha} \,, \quad \text{and so } \Delta k_\alpha = \frac{\pi}{L_\alpha} \,. \tag{16.22}$$

Then

$$\text{cell volume} = \Delta k_x \, \Delta k_y \, \Delta k_z = \frac{\pi^3}{L_x L_y L_z} = \frac{\pi^3}{V} \,, \tag{16.23}$$

and (16.20) becomes

$$\Omega_{\text{tot}} = 2 \times \frac{\frac{1}{8} \times \frac{4}{3}\pi \left(\frac{2\pi f}{c}\right)^3}{\pi^3/V} = \frac{8\pi f^3 V}{3c^3} \,, \tag{16.24}$$

just as obtained for condition (a). This is no surprise; in going from condition (a) to (b), we reduced the volume of $\boldsymbol{k}$ space by a factor of 8, but we also reduced the volume of a cell by the same factor, so Ω_{tot} is unchanged.

Both conditions (a) and (b) give the same number of states Ω_{tot}, so we conclude that the density of states is

$$g(f) = \Omega'_{\text{tot}}(f) = \frac{8\pi f^2 V}{c^3} \,. \qquad (16.25)$$

A comment on this calculation of $g(f)$ The fact that conditions (a) and (b)—and other similar conditions that can be used—give the same number of states indicates that there might be something simple hiding behind our analysis. Although each frequency in a continuum can have only infinitesimal energy, there are an infinite number of such waves in any cell in $\boldsymbol{k}$ space. It seems that in our idealised oven, perhaps all of these waves' energies integrate to give the same energy as would be contained in just one standing wave occupying a cell—although granted we *have* postulated a continuum of waves in the oven. Even so, we did assume on page 66 that at thermal equilibrium the total energy in one frequency state is determined by the wall oscillators, and nothing was said about whether this energy could be held by only a continuum of waves, or whether standing waves could also possess it.

Textbooks usually choose one or both of conditions (a) and (b) above: they tend to treat the oven as full of resonating waves, even though we might presume that an ideal oven would be made of perfectly black material whose walls would not reflect waves at all, and so would not allow the wave to resonate by bouncing back and forth. But we see here that this assumption of resonance is not actually necessary. In analogy, we saw on page 20 that there is more than one way to define the state of an ideal-gas particle via a cell in phase space by using any multiple of Planck's constant, but that luckily what results is a unique expression for *increases* in entropy. The difficulty in defining the state of the gas particle is due to our insisting on *counting* states, which entails the notion of a discrete state.

The same ideas of counting apply to the waves in the oven. Defining and then counting their states by constructing discrete cells in wavenumber space can be problematic—do we use wavelengths or half wavelengths, and why does it not matter what we do? But in the end we have some kind of counting procedure that gives a seemingly unique expression for $g(f)$, and this expression turns out to produce the experimentally verified expression for the spectral energy density $\varrho(f)$. It's certainly interesting and nontrivial why this should be so. Ideas of counting frequency states are intimately related to the quantum mechanical idea of representing the waves by a "gas" of photons. A fundamental difference between photons and the particles of our ideal gas in Section 3.1 is that the number of photons in an oven is not constant over time, unlike the number of ideal gas particles in a container. More discussion of the idea of photons comprising such a gas can be found in [4].

An Alternative Approach to the Maths of Counting States   On page 53 we highlighted the fact that calculating the density of states g by differentiating Ω_{tot} is essentially the same as calculating $\mathrm{d}\Omega_{\text{tot}}$. Whereas Ω_{tot} requires the calculation of the volume of a sphere in $\boldsymbol{k}$ space, $\mathrm{d}\Omega_{\text{tot}}$ requires the calculation of that sphere's surface area.

The same idea applies here. Our calculations of $g(f)$ in conditions (a) and (b) above were really no different to counting the number of cells within a thin spherical shell of radius k, which could be done by treating the volume as surface area times shell thickness dk. In essence we calculated the area of the spherical shell by differentiating its volume $4\!/3\,\pi k^3$ with respect to k to arrive at $4\pi k^2$. In the same way, textbooks usually calculate $g(f)$ by considering this shell of $\boldsymbol{k}$ space. They do this for e.g. condition (a) by writing

$$g(f)\,df = 2 \times \text{number of cells in spherical shell of radius } k \text{ and thickness } dk$$

$$= 2 \times \frac{\text{volume of spherical shell of radius } k \text{ and thickness } dk}{\text{volume of one cell}}$$

$$= \frac{2 \times 4\pi k^2 \, dk}{8\pi^3/V} \;=\; \frac{2 \times 4\pi\left(\frac{2\pi f}{c}\right)^2 \frac{2\pi\,df}{c}}{8\pi^3/V} \;=\; \frac{8\pi f^2 V\,df}{c^3}\,, \tag{16.26}$$

which is (16.25) again without any mention of Ω_{tot} and so without having to differentiate it with respect to f. This is fine. We only calculated Ω_{tot} to emulate and reinforce our approach in Section 3.1 for calculating the number of energy states of an ideal gas.

The End Product: Planck's Law

Now that we have $\varepsilon(f)$ in (16.6) and $g(f)$ in (16.25), we can go back to (16.2) to write *Planck's law*:

$$\boxed{\varrho(f) = \frac{8\pi h f^3/c^3}{e^{\frac{hf}{kT}} - 1}\,.} \tag{16.27}$$

Compare Planck's result for the energy density with the prior result of Rayleigh and Jeans, who used the classical expression $\varepsilon(f) = kT$, based on equipartition with 2 d.o.f., as can be seen in (16.7). This gave them an energy density of

$$\varrho_{RJ}(f) = \frac{8\pi f^2 kT}{c^3}\,. \tag{16.28}$$

This equation's very wrong prediction of arbitrarily large amounts of radiation at high frequencies was called the "ultraviolet catastrophe". Planck's result rested on energy quantisation, and marked the beginning of quantum theory. Note that Planck's expression reduces to that of Rayleigh and Jeans in the low-frequency limit.

16.2 Total Energy per Unit Volume of the Oven, U

The total radiation energy in a unit volume is the spectral energy density integrated over all frequencies:

$$U = \int_0^\infty \varrho(f)df\,. \tag{16.29}$$

Writing $x \equiv \frac{hf}{kT}$ converts this to

$$U = \frac{8\pi k^4 T^4}{c^3 h^3} \underbrace{\int_0^\infty \frac{x^3 dx}{e^x - 1}}_{=\,\pi^4/15} \;=\; \underbrace{\frac{8\pi^5 k^4}{15 c^3 h^3}}_{=\,4\sigma/c,\text{ where } \sigma \text{ is the Stefan–Boltzmann}} T^4. \tag{16.30}$$

$$\text{constant, defined in (16.40) ahead.}$$

That is, the total photon energy inside an oven of volume V and temperature T is

$$\boxed{\text{total photon energy inside oven} = V\,\frac{4\sigma}{c}\,T^4\,,}$$

(16.31)

where $\sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the *Stefan–Boltzmann constant*, defined in (16.40) ahead.

16.3 Planck's Law in terms of Wavelength

Since an increase in frequency $\mathrm{d}f$ is accompanied by a *decrease* in wavelength $-\mathrm{d}\lambda$, where $\mathrm{d}f$ and $-\mathrm{d}\lambda$ have the same sign, define the energy density over wavelength $\varrho(\lambda)$ by

$$\varrho(\lambda)(-\mathrm{d}\lambda) \equiv \varrho(f)\mathrm{d}f\,,$$

(16.32)

which leads to

$$\varrho(\lambda) = \frac{8\pi hc/\lambda^5}{e^{\frac{hc}{\lambda kT}} - 1}\,,$$

(16.33)

where we have used $f = c/\lambda$ and $\mathrm{d}f/\mathrm{d}\lambda = -c/\lambda^2$. Although we won't do so here, it's not difficult to use (16.33) to derive *Wien's law*, which gives the "most copiously emitted wavelength" λ_0. We simply solve $\varrho'(\lambda_0) = 0$ numerically to get

$$\boxed{\lambda_0 = \frac{\text{constant}}{T} \simeq \frac{2.898\,\mathrm{mm\,K}}{T}\,.}$$

(16.34)

The same idea serves to determine the "most copiously emitted frequency" f_0 from (16.27). The result is

$$\frac{f_0}{T} = \text{constant} \simeq 58.8\,\mathrm{GHz\,K^{-1}}.$$

(16.35)

It might be thought that $f_0\lambda_0 = c$, but such is not the case! The reason for this apparent anomaly is that the phrase "most copiously emitted frequency" implies that there are various frequencies present, like balls of various colours, and we are finding the ball of the most common colour—and similarly for wavelength. But this is not quite the case; we have assumed that frequency and wavelength are *continuous*, so that the equal-width frequency bins that we are essentially comparing to find the "most copiously emitted frequency" do not map to equal-width wavelength bins, since frequency and wavelength are not related linearly. So the phrase "most copiously emitted" should be taken with a grain of salt.

16.4 Radiation Exiting the Oven

Make a small hole in the oven. How much energy escapes per second? To determine this, place the origin of a cartesian coordinate system at the hole, and let the wall containing the hole be the xz plane, with the y axis pointing into the oven. The energy passing through the hole of area $\mathrm{d}A$ in a time Δt is that of *some* of the photons that are within a distance $c\Delta t$ of the hole: the photons of interest are those moving in the correct direction to encounter the hole. The energy escaping the hole from a volume $\mathrm{d}V$ at a distance r from the hole is then (call it α) determined by the solid angle subtended by $\mathrm{d}A$ as seen

from dV, or, equivalently, the area that dA projects onto a sphere of radius r centred at dV:

$$\alpha = U\,dV \times \frac{\text{``area on sphere'' subtended by } dA \text{ as seen from } dV}{4\pi r^2}. \tag{16.36}$$

The "area on the sphere" is the dot product of the hole area expressed as a vector, $-dA\,\boldsymbol{e}_y$ (where $\boldsymbol{e}_y$ is the usual y basis vector, of unit length), with the "look direction" from dV, which is minus the radial basis vector, $-\boldsymbol{e}_r$ (of unit length):

$$\begin{aligned}\text{``area on sphere''} &= -dA\,\boldsymbol{e}_y \cdot (-\boldsymbol{e}_r) = dA\,(0,1,0) \cdot (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\\ &= dA\,\sin\theta\sin\phi\,, \end{aligned} \tag{16.37}$$

where θ, ϕ are the usual spherical polar coordinates. So (16.36) becomes

$$\alpha = U\,r^2 \sin\theta\,dr\,d\theta\,d\phi \times \frac{dA\,\sin\theta\sin\phi}{4\pi r^2}\,. \tag{16.38}$$

That means the total energy passing through the hole of area dA in a time Δt is

$$\begin{aligned}\int \alpha &= \int_0^\pi d\theta\,\sin^2\theta \int_0^\pi d\phi\,\sin\phi\,\frac{U\,dA}{4\pi}\\ &= \frac{Uc}{4}\,dA\,\Delta t\,. \end{aligned} \tag{16.39}$$

So the energy radiated per unit hole area per unit time is $Uc/4$. In other words,

$$\text{power radiated per unit hole area} = \frac{Uc}{4} \overset{(16.30)}{=\!=\!=\!=} \frac{2\pi^5 k^4}{15c^2 h^3}\,T^4 \equiv \sigma T^4\,, \tag{16.40}$$

where $\sigma \simeq 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the *Stefan–Boltzmann constant.*

Note that although the U here is the total energy per unit volume of the oven, it can also stand for the total energy per unit volume of the oven per unit frequency, or per unit wavelength; i.e., $\varrho(f)$ or $\varrho(\lambda)$. The conversion to a power was simply accomplished with the factor of $c/4$. For example, the power radiated per unit hole area per unit wavelength is, from (16.33),

$$\frac{\varrho(\lambda)c}{4} = \frac{2\pi hc^2/\lambda^5}{e^{\frac{hc}{\lambda kT}} - 1}\,. \tag{16.41}$$

16.5 Radiation from a Black Body (usually called "blackbody radiation")

We now return to where we started, with the task of calculating how much radiation is emitted by a black body. We argued that it must emit what it absorbs, which is thus the same (Planck) spectrum produced by the oven. So we infer that the black body's radiated power equals that which emerges from a hole made in the side of the oven. That means

the power radiated by a black body per unit area of its surface is σT^4, from (16.40). If it has area A, then

$$\text{total power emitted by black body} = A\sigma T^4. \tag{16.42}$$

A body that isn't black has an *emissivity* $e(\lambda, T)$, typically measured experimentally. The emissivity is sometimes approximated by a constant e, so

$$\boxed{\text{total power emitted by any body} \simeq Ae\sigma T^4.} \tag{16.43}$$

Example: The sun's power output is fitted well by a Planck spectrum for $T = 5800\,\text{K}$, so we can treat it as a black body with this surface temperature. It has radius $r = 700{,}000\,\text{km}$, and its internal temperature (averaged over its volume) is about $10^7\,\text{K}$. If the sun's thermonuclear reactions stopped today, how long could it continue to emit at its current rate?

Use (16.31) and (16.43) to write

$$\text{this time period} = \frac{\text{total photon energy inside}}{\text{total current luminosity}}$$

$$= \frac{\frac{4}{3}\pi r^3 \frac{4\sigma}{c} T_{\text{internal}}^4}{4\pi r^2 \, \sigma \, T_{\text{surface}}^4} = \frac{4r}{3c}\left(\frac{T_{\text{internal}}}{T_{\text{surface}}}\right)^4$$

$$= \frac{4 \times 7\underline{\underline{8}}}{3 \times 3\underline{\underline{8}}} \times \left(\frac{10^7}{5800}\right)^4 \times \frac{1}{31.5\underline{\underline{6}}} \text{ years}$$

$$\simeq 870{,}000 \text{ years.} \qquad \underline{\text{Answer}} \tag{16.44}$$

16.6   The Greenhouse Effect

Consider a glass greenhouse with its ceiling somewhere above the ground, along with an incoming flux density J_i (i.e. power per unit area) of solar radiation. Almost all of the mainly visible light of the solar spectrum passes through glass without being absorbed and re-scattered. It heats the ground—but not to solar temperatures of course, so the ground re-radiates an "outgoing" flux density J_o, but at a much longer wavelength: typically largely in the infra-red.

Glass absorbs some of this infra-red light and re-radiates a portion αJ_o back to the ground, which further heats the ground. When equilibrium is reached, there can be no net flow anywhere, so consider an imaginary plane between glass and ground. The flux density down through this imaginary plane, $J_i + \alpha J_o$, equals the flux density up through it, J_o, so

$$J_o = \frac{J_i}{1-\alpha}. \tag{16.45}$$

That means $J_o > J_i$. Alternatively, place the imaginary plane above the glass. Now what comes down through *this* imaginary plane, J_i, equals what goes up through it, $(1-\alpha)J_o$, giving us (16.45) again.

Now consider J_o with and without the glass ceiling present:

$$\frac{\sigma T^4(\text{glass})}{\sigma T^4(\text{no glass})} = \frac{J_o(\text{glass})}{J_o(\text{no glass})} = \frac{\frac{J_i}{1-\alpha}}{J_i} = \frac{1}{1-\alpha}. \tag{16.46}$$

Therefore the glass heats the ground in the ratio

$$\frac{T(\text{glass})}{T(\text{no glass})} = \left(\frac{1}{1-\alpha}\right)^{1/4} > 1. \tag{16.47}$$

Example: Earth receives a mean flux density of solar energy of $J_i = 175 \text{ Wm}^{-2}$, averaged over all latitudes and all times of the day and night. Of this, 90% is absorbed and 10% reflected back into space. Assuming Earth's infra-red emissivity is 0.9, what would be the average temperature on Earth's surface if it had no atmosphere?

90% of 175 Wm^{-2} is 158 Wm^{-2}. At equilibrium this must all be re-radiated, so

$$0.9 \, \sigma T^4 = 158 \text{ Wm}^{-2}. \tag{16.48}$$

Thus

$$T = \left(\frac{158}{0.9 \times 5.67^{-8}}\right)^{1/4} \text{K} = 236 \text{ K} = -37^{\circ}\text{C}. \qquad \underline{\underline{\text{Answer}}} \tag{16.49}$$

Now introduce our atmosphere, which absorbs and re-radiates almost all of the radiation leaving the ground. What temperature results?

In this case $\alpha = 1/2$, since we are taking all of J_o to be absorbed, and half of this is radiated back to the ground. So (16.47) gives

$$\frac{T(\text{atmos.})}{T(\text{no atmos.})} = \left(\frac{1}{1-\alpha}\right)^{1/4} = 2^{1/4}, \tag{16.50}$$

resulting in

$$T(\text{atmos.}) = 2^{1/4} \times 236 \text{ K} = 281 \text{ K} = 8^{\circ}\text{C}. \qquad \underline{\underline{\text{Answer}}} \tag{16.51}$$

16.7 Thermal Noise and Maximum Channel Capacity

Thermal fluctuations in electrical circuits produce noise that manifests as voltage fluctuations. To explore this, model a circuit resistor as a one-dimensional oven of length L, carrying electromagnetic waves as before. We will mimic the derivation of $\varrho(f)$ of Section 16.1, but it will be simpler this time. We'll also focus not on $\varrho(f)$ by itself, but on the total energy inside the resistor over a frequency range $f \to f + \mathrm{d}f$, which is $\varrho(f)\,\mathrm{d}f\,L$.

The one-dimensional analogue of (16.2) is

$$\varrho(f) = \frac{\varepsilon(f)\,g(f)}{L}. \tag{16.52}$$

As ever, $\varepsilon(f)$ is the mean thermal energy of an oscillator, now inside the resistor instead of an oven wall. This is again (16.6). Usually we are concerned with sub-GHz frequencies at room temperature, for which $hf \ll kT$, so that $\varepsilon(f) \simeq kT$ (the equipartition value).

The discussion beginning on page 67 of how to calculate the density of states $g(f)$ applies equally here, but now the wave vector space is one dimensional. We can choose either of conditions (a) or (b) on pages 68 and 69, and choose to calculate Ω_{tot} or use the alternative approach that produced (16.26). The results will all be the same.

We'll choose condition (a) and calculate Ω_{tot}. A continuum of waves is held in the resistor (so that both signs of k are used), and a whole number n of wavelengths fit into the resistor's length L. Again remembering there are 2 polarisations,

$$\Omega_{\text{tot}} = 2 \times \text{number of cells}$$

$$= 2 \times \frac{\text{length of interval } [-k, k]}{\text{length of one cell } (= \Delta k)} = \frac{4k}{\Delta k}. \tag{16.53}$$

But $n\lambda = L$, so

$$k = \frac{2\pi}{\lambda} = \frac{2\pi n}{L}, \quad \text{and therefore } \Delta k = \frac{2\pi}{L}. \tag{16.54}$$

Also $k = 2\pi f/c$, so the number of frequency states is

$$\Omega_{\text{tot}} = \frac{4k}{\Delta k} = \frac{4 \times \frac{2\pi f}{c} L}{2\pi} = \frac{4fL}{c}. \tag{16.55}$$

Thus

$$g(f) = \Omega'_{\text{tot}}(f) = \frac{4L}{c}. \tag{16.56}$$

Now use (16.52) to get the total energy inside the resistor over a frequency range $f \to f + \mathrm{d}f$ for the sub-GHz frequencies of interest:

$$\text{total energy within resistor in } \mathrm{d}f = \varrho(f)\,\mathrm{d}f\,L = \varepsilon(f)\,g(f)\,\mathrm{d}f \simeq kT\,4L\,\mathrm{d}f/c. \tag{16.57}$$

Typically we require the total energy held in some frequency range B, where B is called the *bandwidth*. Integrate (16.57) to find this total energy:

$$\text{total energy within resistor in } B \simeq 4kTLB/c. \tag{16.58}$$

If this energy all emerges in a time L/c by moving along the resistor at speed c (an adequate approximation), then

$$\text{power out} = \frac{\text{energy out}}{\text{time taken}} = \frac{4kTLB/c}{L/c} = 4kTB. \tag{16.59}$$

This is an average of course; it's all based on the idea of electromagnetic fluctuations occurring inside the resistor. It is given the name *Nyquist's theorem for thermal noise in circuits*:

$$\boxed{\text{noise generated in a circuit} = 4kTB.} \tag{16.60}$$

This noise power manifests as a fluctuating voltage, since the power dissipated in a resistor R due to a voltage V across it is V^2/R. In that case

$$\langle V^2/R \rangle = 4kTB, \tag{16.61}$$

so that the mean-square voltage arising from the noise is

$$\langle V^2 \rangle = 4R\,kTB. \tag{16.62}$$

The noise of complicated circuits arises from many sources interacting with each other in various ways, so that more generally the "4" in Nyquist's theorem (16.60) is replaced by the circuit's *noise factor*, F.

Why might we be interested in having some particular bandwidth B? One reason in an era of communication is the *Shannon–Hartley theorem*, which states that the maximum transmission capacity C that some data-transmitting channel can have, below which we can always arrange for an arbitrarily low error rate, is a function of the bandwidth B that we use, and the signal-to-noise ratio S/N that we require:

$$\underbrace{C}_{\text{bits/unit time}} = B \log_2(1 + S/N) , \qquad (16.63)$$

where a "bit" is a binary digit: 0 or 1, and S, N are the signal and noise powers respectively. For example, if we want to send a signal with signal-to-noise ratio of $S/N = 10$, and we have $B = 1$ MHz of bandwidth at our disposal, then the maximum throughput for which we can ever hope to arrange an arbitrarily low error rate is 1 MHz $\times \log_2 11$, or 3.46 megabits per second.

> By "hoping to arrange an arbitrarily low error rate" we mean the following. When signals are sent down a line, errors can always be introduced by noise en route and in the receiver. Sophisticated error-correction algorithms can correct some of these errors; but the higher the percentage of errors we wish to correct, the more sophisticated our algorithm will need to be. The Shannon–Hartley theorem puts an upper bound on the amount of information that we can ever send, even if we have an all-powerful error-correction algorithm corrects 100% of the errors.

Equation (16.63) shows that to achieve a high transmission rate, we should have a large bandwidth.

> This makes sense from the viewpoint of Fourier analysis: having a large bandwidth means we have a large range of frequencies at our disposal, and that means we're able to craft signal waveforms with "tighter turns" in them. So for example, if we are transmitting a square wave that encodes binary digits, more bandwidth allows us to squeeze more oscillations of the wave into a given length. After all, signals travel at a set speed, so if we want to send more of them per second, we have to make each one shorter. This need for a "broader band" of frequencies to send more data is the origin of the term *broadband* used frequently by Internet service providers.

But in an electronic circuit, the noise N is given by $F\,kTB$. So increasing the bandwidth has two competing effects: it partly acts to increase the maximum transmission capacity C, but it also partly acts to *decrease* C because increasing B introduces more noise into the system, through Nyquist's theorem. You can see that calculating a channel's maximum capacity requires knowledge of its noise factor and temperature.

17   Theory of Electric Conduction

Here we outline an example of where the classical picture of electric conduction fails, and investigate how quantum mechanics steps in to begin to solve the problem.

17.1   The Classical Picture

The electrons that carry a current through a wire of end area A can be modelled as a gas of free electrons inside a metal lattice.

Write

$$n = \text{electron number density}, \qquad q = \text{electric charge},$$
$$v_d = \text{electron drift speed}, \qquad A = \text{area of wire},$$
$$m = \text{electron mass}. \qquad\qquad\qquad\qquad\qquad\qquad (17.1)$$

The charge crossing A in a time Δt equals the charge contained in the swept volume $Av_d\Delta t$, which is $nqAv_d\Delta t$. The electric current in the wire is then

$$I = \frac{\text{charge passed}}{\Delta t} = \frac{nqAv_d\Delta t}{\Delta t} = nqAv_d, \qquad (17.2)$$

and the current density is

$$J = \frac{I}{A} = nqv_d. \qquad (17.3)$$

On average, v_d is about half the speed picked up from an acceleration due to a force $\mathcal{E}q$ (exerted by the electric field $\mathcal{E}$) that acts for a time of $\lambda/\bar{v}$, where λ is the mean free path of the electrons, and $\bar{v}$ is their mean thermal speed (from the Maxwell speed distribution). So

$$v_d = \frac{1}{2}\frac{\mathcal{E}q}{m}\frac{\lambda}{\bar{v}}, \qquad (17.4)$$

and therefore

$$J = \frac{n\mathcal{E}q^2\lambda}{2m\bar{v}}. \qquad (17.5)$$

What does experiment say? Ohm's rule gives the resistance across a length ℓ as

$$R = \frac{V}{I} = \frac{\mathcal{E}\ell}{JA} \equiv \frac{\varrho\ell}{A}, \qquad (17.6)$$

where ϱ is the *resistivity* of the conductor. This means that, experimentally,

$$J = \mathcal{E}/\varrho. \qquad (17.7)$$

This agrees with (17.5), which predicts $J \propto \mathcal{E}$. So far so good, for our (classical) microscopic model of electric current. But notice that (17.5), (17.7) also imply

$$\varrho = \frac{\mathcal{E}}{J} = \frac{2m\bar{v}}{nq^2\lambda}. \qquad (17.8)$$

We know that $\bar{v} \propto \sqrt{T}$, from the Maxwell speed distribution. We might use (13.2) to write the mean free path amongst fixed lattice atoms, with ν of these atoms per unit volume and each with cross section σ:

$$\lambda = \frac{1}{\nu\sigma}. \qquad (17.9)$$

In that case, we conclude that $\varrho \propto \sqrt{T}$. However, this is experimentally wrong. Experiments show that $\varrho \propto T$. We can try to fix this by modifying the mean free path of the electrons among the lattice atoms. Suppose an electron sees a lattice atom (mass M) vibrating in three dimensions with circular frequency ω and amplitude A. Equipartition gives this atom a vibrational energy

$$\tfrac{3}{2}M\omega^2 A^2 = \tfrac{3}{2}kT \qquad (17.10)$$

so that it presents a cross section to wandering electrons of

$$\sigma = \pi A^2 = \frac{\pi k T}{M \omega^2} \,. \tag{17.11}$$

This combines with (17.9) to give

$$\lambda = \frac{M \omega^2}{\nu \pi k T} \,. \tag{17.12}$$

With this temperature-dependent λ, (17.8) predicts $\varrho \propto T^{3/2}$. But this still disagrees with experiment.

17.2   The Quantum Picture

Quantum mechanics clears the above puzzling aspects of conduction in great depth by incorporating the proposition that electrons are fermions. This proposition can be taken as confirmed, thanks to the experimental success of the predictions that result.

In Section 15 we saw that at most only one fermion can occupy a given quantum state. This is a severe restriction that drastically alters the quantum behaviour of electrons as compared to a classical treatment. But it successfully predicts properties of conductors, semi-conductors, and insulators that are not predicted classically. We'll see some of this in what follows.

The simplest quantum viewpoint models the gas of electrons as noninteracting particles in a cubic infinite square well of side length L. Solving Schrödinger's equation for such a "particle in a box" is a standard and straightforward exercise in introductory quantum mechanics textbooks. The infinite potential energy at the walls of the box constrain each particle to the box. Inside, the zero potential energy makes Schrödinger's equation easy to solve for its wave functions and energy eigenvalues. These eigenvalues are interpreted as the quantised energy levels of the gas of particles, and are

$$E_{n_x n_y n_z} = \underbrace{\frac{h^2}{8mL^2}}_{\equiv\, E_1} \left(n_x^2 + n_y^2 + n_z^2 \right). \tag{17.13}$$

How many electrons can fill these quantum states up to energy E? This number of electrons equals the number of states Ω_{tot} with energy $\leqslant E$ (recall Section 3.1), which—with two spins allowed—is twice the number of unit cubes in $n_x n_y n_z$ space, in the octant of radius $\sqrt{n_x^2 + n_y^2 + n_z^2}$, i.e. of radius $\sqrt{E/E_1}$. This number is twice the volume of the octant of positive n_x, n_y, n_z out to this radius, or

$$\text{number of electrons} = \Omega_{\text{tot}} = 2 \times \frac{1}{8} \times \frac{4\pi}{3} \left(\sqrt{\frac{E}{E_1}} \right)^3 = \frac{\pi}{3} \left(\frac{E}{E_1} \right)^{3/2} = nL^3 \,, \tag{17.14}$$

since there are n electrons per unit volume in the cube. At $T = 0$ these will fill the lowest energy states up to the *Fermi energy* E_F, which in this context is just another name for the chemical potential μ at this temperature: the reason is because a plot of (15.10)

at $T = 0$ shows that the occupation number changes abruptly from 1 to 0 at energy μ. So we conclude that at $T = 0$,

$$E_F = \frac{h^2}{8m} \left(\frac{3n}{\pi} \right)^{2/3}. \qquad (17.15)$$

Example: Calculate E_F ($T = 0$) for copper metal, which has $n = 8.47 \times 10^{28}$ electrons/m^3. The electron's mass is $m = 9.11 \times 10^{-31}$ kg.

Equation (17.15) gives, in SI units with a final conversion to electron volts,

$$E_F = \frac{\left(6.626^{-34}\right)^2}{8 \times 9.11^{-31}} \left(\frac{3 \times 8.47^{28}}{\pi} \right)^{2/3} \frac{1}{1.6^{-19}} \text{ eV} \simeq 7.0 \text{ eV.} \qquad \underline{\text{Answer}} \qquad (17.16)$$

– Note that the gain in energy from lattice atoms $\simeq kT$ ($\simeq 1/40$ eV at room temperature), so only those electrons within kT of E_F can be excited into unoccupied levels. (Incidentally, the characteristic width of the sloping section of the $\bar{n}(E)$-vs-E plot for fermions is around kT as well.)

– The electrons don't scatter off each other. The reason is because if this were to happen, pushing one into an unoccupied level, the other would have to drop to a lower unoccupied level—which doesn't exist, because the levels are all filled. So these electrons don't obey the Equipartition Theorem.

Mean Electron Energy

This is

$$\bar{E} = \frac{\text{total e}^- \text{ energy}}{\text{total no. of e}^-} = \frac{1}{nL^3} \int_0^{nL^3} E \, d\Omega_{\text{tot}} , \qquad (17.17)$$

where Ω_{tot} is the number of electrons out to some energy E. Equation (17.14) gives $d\Omega_{\text{tot}} = \pi/2 \, E_1^{-3/2} E^{1/2}$. In that case,

$$\bar{E} = 3/5 \, E_F . \qquad (17.18)$$

Fermi Temperature T_F

Define this via

$$kT_F \equiv E_F . \qquad (17.19)$$

If $T \ll T_F$, the mean energy of the lattice atoms ($\simeq kT$) is $\ll E_F$, so the electron occupation number distribution varies little from its $T = 0$ shape. For copper,

$$T_F \simeq \frac{7.0 \times 1.6^{-19}}{1.38^{-23}} \text{ K} \simeq 81,000 \text{ K} , \qquad (17.20)$$

and we conclude that copper's gas of conduction electrons certainly cannot be treated classically for any reasonable temperature at all.

Fermi Speed v_F

Define this via

$$\tfrac{1}{2}\,mv_F^2 \equiv E_F\,. \tag{17.21}$$

For copper,

$$v_F = \sqrt{\frac{2E_F}{m}} \simeq \sqrt{\frac{2\times 7.0\times 1.6^{\underline{\underline{-19}}}}{9.11^{\underline{\underline{-31}}}}} \ \text{m/s} \simeq 1600\ \text{km/s}\,. \tag{17.22}$$

This is not overly affected by temperature, at least for $T \ll 81{,}000$ K... Compare this to the Maxwell mean speed $\bar{v} \propto T^{1/2}$. At room temperature,

$$\bar{v} = \sqrt{\frac{2\times 1.38^{\underline{\underline{-23}}}\times 300}{9.11^{\underline{\underline{-31}}}}} \ \text{m/s} \simeq 95\ \text{km/s}\,. \tag{17.23}$$

(Note that some people define the Fermi speed via $\tfrac{1}{2}\,mv_F^2 \equiv \tfrac{3}{5}\,E_F$.)

The quantum mechanical point of view considers it more appropriate to replace the $\bar{v}$ in (17.8) with v_F. Since v_F is effectively independent of T, and (17.12) gives $\lambda \propto 1/T$, (17.8) becomes

$$\varrho = \frac{2mv_F}{nq^2\lambda} \propto T\,, \quad \text{as observed experimentally.} \tag{17.24}$$

So quantum mechanics has come to the rescue by predicting the correct dependence on temperature for the electrical resistivity.

17.3 Band Theory of Solids

To understand the degree of availability of free electrons, we must consider the effect of the crystal lattice on the electron energy levels. Solving the Schrödinger equation for an electron in an atom yields a discrete set of energy levels. However, when we bring two atoms close together, the energy of each level changes due to the influence of the other atom. (We can also arrive at this by solving the Schrödinger equation for electrons moving in a periodic potential.)

If we bring N atoms together in a lattice, a particular energy level splits into N levels, forming a band. The band width is determined by the atomic spacing—not by N, so for large N the band is composed of an almost continuous spread of energy levels. Bands are typically a few eV thick, and may overlap.

One piece of direct evidence for bands in solids comes from X ray spectra. Gaseous sodium shows the expected sharp peaks due to energy level quantisation, whereas the same peaks produced from solid sodium are broadened due to the bands being present.

Allowed bands are continuous bands of energy levels for electrons.

Forbidden bands are regions where there are no energy levels.

The band containing the outer electrons is called either the *valence band* if it's full of electrons—i.e. if all of its energy levels are occupied; or the *conduction band* if it isn't full. Valence bands correspond to insulators and semiconductors. Conduction bands correspond to conductors.

We treat materials as well and truly insulators when the width of the first forbidden band above their valence band is > 2 eV. (E.g. the width for diamond is 7 eV.) If the width is < 2 eV, we call the material a semiconductor. Examples are silicon (1.1 eV) and germanium (0.7 eV).

17.4   Insulators and Semiconductors

Consider a material whose (filled) valence band extends to energy E_v. There is then a gap of width E_g which forms a forbidden band, and then the (almost empty) conduction band begins at energy E_c. By symmetry, the Fermi energy E_F lies in about the middle of the gap, so that $E_F = E_v + E_g/2$. What is the number density n_e of electrons in the conduction band? Set $\beta \equiv 1/(kT)$ to write

$$n_e = \frac{1}{L^3} \int_{E_c}^{\infty} \bar{n}(E) g(E) \, \mathrm{d}E = \frac{1}{L^3} \int_{E_c}^{\infty} \frac{g(E) \, \mathrm{d}E}{e^{\beta(E-E_F)} + 1}$$
$$\simeq \frac{1}{L^3} \int_{E_c}^{\infty} e^{-\beta(E-E_F)} g(E) \, \mathrm{d}E \,. \qquad (17.25)$$

The density of states $g(E)$ depends heavily on the material. Remember from Section 3.1 that $g(E) \, \mathrm{d}E = \mathrm{d}\Omega_{\text{tot}}$. For the simple model of particles in a box, Ω_{tot} is given by (17.14):

$$\Omega_{\text{tot}} = \frac{\pi}{3} \left(\frac{E}{E_1} \right)^{3/2}, \quad \text{so } g(E) = \frac{\pi}{2} \frac{E^{1/2}}{E_1^{3/2}} \,. \qquad (17.26)$$

What is more usual is to model the possibly complicated density of states by

$$g(E) = \frac{\pi}{2} \frac{(E - E_c)^{1/2}}{E_1^{3/2}} \,. \qquad (17.27)$$

Also, the electron mass m (inside E_1) is replaced by some effective mass m^* that depends on the nature of the lattice. In this new model, the electron number density of (17.25) becomes

$$n_e \simeq \frac{1}{L^3} \int_{E_c}^{\infty} e^{-\beta(E-E_F)} \frac{\pi}{2} \frac{(E - E_c)^{1/2}}{E_1^{3/2}} \, \mathrm{d}E \,. \qquad (17.28)$$

Write $E - E_F = E - E_c + E_g/2$ and use a change of variables $u \equiv (E - E_c)^{1/2}$. The integral is then straightforward, and gives

$$\frac{n_e}{n} \approx \left(\frac{kT}{E_F} \right)^{3/2} \exp \frac{-E_g}{2kT} \,. \qquad (17.29)$$

Example: Calculate n_e at room temperature using the electron number density and Fermi energy of copper, for the two cases of forbidden band widths 7 eV (an insulator) and 1 eV (a semiconductor).

We use $kT = 1/40$ eV, $n = 8.47 \times 10^{28}$ m^{-3}, and $E_F = 7$ eV:

Insulator: $E_g = 7$ eV, so

$$n_e \approx 8.47^{28} \left(\frac{0.025}{7} \right)^{3/2} \exp \frac{-7}{0.05} \approx 3 \times 10^{-36}. \qquad (17.30)$$

Semiconductor: $E_g = 1$ eV, so

$$n_e \approx 8.47 \underset{\underline{28}}{} \left(\frac{0.025}{7} \right)^{3/2} \exp \frac{-1}{0.05} \approx 4 \times 10^{16}. \tag{17.31}$$

The comparatively huge number of conduction electrons in the semiconductor is evident. Their number is also affected by temperature: materials that are insulators as $T \to 0$ can become semiconductors as the temperature rises.

17.5   Diodes

Electrons that jump across the gap into the conduction band leave behind a "hole" in the valence band. This can be treated as another particle, but with positive charge, that contributes to the current.

By adding impurities—small amounts of other elements—to the semiconductor, we can cause more or less free electrons to move within the lattice (actually orbiting say an arsenic atom at a large distance), or holes to move within the lattice (actually orbiting say a gallium atom at a large distance). These donor impurities create electron energy levels which allow electrons to be excited; thus they cause semiconductors to conduct much better than if the conduction were due to thermal excitation alone.

An *n-type* semiconductor is one that has been "doped" with an element such as arsenic or antimony, elements that donate electrons (negative change—hence the name n-type). These donated electrons populate new energy levels that appear at the *top* of the forbidden band. Thermal excitation or an external electric field can quite easily bump these electrons into the conduction band; hence the dramatic increase in conduction due to the impurity.

A *p-type* semiconductor is doped with e.g. gallium or indium: metals that accept electrons or, equivalently, donate holes: positive change—hence the name p-type. These holes populate new energy levels that appear at the *bottom* of the forbidden band. Thermal excitation or an external electric field can now easily bump electrons from the top of the valence band into these holes.

Suppose we have two doped semiconductors: the p-type has holes that are free to wander about its lattice, and the n-type has electrons that are free to wander. If we join them together to form a *pn-semiconductor*, some of the free electrons in the n-type close to the junction will move to fill the immediately adjacent holes on the p-type side of the junction. This creates a slight excess of negative charge on the p-type side of the junction, and a slight excess of positive charge on the n-type side of the junction. A permanent internal electric field has now been created across the junction, pointing from the slightly positive n-type side to the slightly negative p-type side.

There are now two processes continuously occurring across the junction:

Thermal: Free electrons from the n-type side are always being propelled by thermal fluctuations to join the excess negative charge on the p-type side. That serves to increase the field's strength. The Boltzmann distribution can be applied to determine how many thermal electrons are propelled across the junction.

Electromagnetic: The strong field in turn keeps returning electrons from the p-type side back home across the junction. That acts to reduce the field's strength, and the whole situation is in a steady state.

The energy term in the Boltzmann distribution (10.20) will be the increase in potential energy that the n-type's free electrons "see" from where they are to the far side of the junction before they are thermally propelled across.

> Note that I am being descriptive in my wording here because the details are easy to get confused. The field across the junction points from n-type to p-type; that means the n-type's free electrons experience a drop in electromagnetic potential when they are forced across the junction by thermal fluctuations. But because they have negative charge, their potential *energy* will *increase* during this process.
>
> Should the energy term in (10.20) be the potential energy seen on the far side of the junction by the electrons, or the potential energy *increase* that they see across the junction? In fact either will work. Remember that potential energy is only ever defined up to an additive constant. We can include this constant in the exponential in (10.20), but it will only get absorbed into the constant of proportionality in that equation. So we might as well set it equal to zero.

So set the potential energy of a free electron on the n-type side of the junction to zero. It then sees a potential energy on the p-type side of $U_0 > 0$.

We can envisage the continuous thermal and electromagnetic flows of electrons as follows. (In the next paragraphs, it helps to consider the current of electrons as a particle current, so we'll use a lower-case i to refer to electron current, and an upper-case I to refer to conventional circuit-theory current.)

Thermal: A current $i_{\text{thermal}} > 0$ of electrons going from n to p (i.e. this is not electric current of the "conventional" sign) due to thermal fluctuations boosts these electrons into a higher potential energy. The number of electrons forming the current obeys the Boltzmann distribution, which closely approximates the tail of the Fermi-Dirac distribution for the electron occupation number at energies above the forbidden band, or $> E_F$.

Electromagnetic: This current of electrons $i_0 > 0$ (again not "conventional" electric current) flows from p to n across the junction, driven by the permanent internal electric field. A typical value of i_0 is about a milliamp.

In equilibrium these flows are balanced, so there is no net current at all.

Suppose that now we apply a *bias voltage*, by connecting the p-type to one terminal of an electric cell of voltage $V_b > 0$, and the n-type to the other terminal. We'll "forward bias" the diode, connecting the p-type to the positive terminal. Now the potential energy of an electron on the p side of the junction decreases from U_0 to $U_0 - eV_b$ (where $e = 1.6 \times 10^{-19}$ C). This lowers the potential barrier for the n-type's free electrons to form the i_{thermal} current, but doesn't affect i_0, which is a kind of ever-present background current due to the base conditions existing internally to the junction. Set

$$I \equiv \text{conventional electric current through diode from p-type to n-type}$$

$$= \text{electron current through diode from n to p}$$

$$= i_{\text{thermal}} - i_0 \,. \tag{17.32}$$

But notice that

$$i_{\text{thermal}} \propto \exp \frac{-\Delta \text{ pot. energy } (\text{n} \to \text{p})}{kT}$$

$$= C \exp \frac{-(U_0 - eV_b)}{kT} = C' e^{\frac{eV_b}{kT}} \tag{17.33}$$

for some normalisations C, C'. For the case of no bias, $i_{\text{thermal}} = C'$ which therefore equals i_0, because $I = 0$ with no bias. That means (17.32) can always be written

$$I = i_0 e^{\frac{eV_b}{kT}} - i_0 = i_0 \left(e^{\frac{eV_b}{kT}} - 1 \right). \tag{17.34}$$

The case of a negative value of V_b is a *reverse bias*, meaning the p-type has been connected to the negative cell terminal (and n-type to positive). Here the potential energy step that the electrons must jump (n→p) gets higher, so $i_{\text{thermal}} \to 0$ and the only current I across the junction is the small background electric current $-i_0$ (the negative sign is due to I being *conventional* electric current). So the diode passes almost no current when it's reverse biased, as (17.34) shows. Commercial diodes can be quite robust when reverse biased, and will only pass 1 or 2 milliamps even when V_b equals minus several hundred volts.

But when the diode is forward biased, the potential energy barrier seen by the electrons drops. Given the exponential nature of the Boltzmann distribution, this allows a huge number of thermally excited electrons to jump up the potential step. So a forward bias produces a current I which is typically several amps for $V_b = +1$ volt, and which in principle can be huge as (17.34) shows; whereas a reverse bias produces essentially no current at all. Diodes thus pass current pretty much in one direction only, which makes them very useful in electronic devices.

18 Final Comments and Acknowledgements

Historically, statistical mechanics resulted from physicists' efforts to put thermodynamics onto a more mathematical footing. One thing I have tried to emphasise in these notes is that this mathematical basis should not be taken as implying that statistical mechanics is a completely "closed" subject, whose concepts are now completely well defined, easy to calculate, and subject to the application of endless rigor.

For example, I put some emphasis on explaining the difficulties involved with counting states in Section 3.1. The states of simple systems can certainly be counted, but in general it seems to be impossible to count the states of more complex systems exactly, even in some idealised way. A good example of this difficulty is the standard expression for the entropy of an ideal gas, not covered in these notes but which is found in some textbooks, such as [5]. What might be surprising is that this entropy expression fails the Third Law of Thermodynamics, because it does not vanish at zero temperature. Its derivation is based on setting the gas's number of states at a given energy E to be what I have called $\Omega_{\text{tot}}(E)$ instead of the more correct $\Omega(E)$, so is headed in the wrong direction from the very start. Given that the counting procedure used to derive that standard expression for the gas's entropy is already an approximation, we can certainly introduce further approximations to make the entropy vanish at zero temperature. But it's not clear whether any such expression for the entropy of an ideal gas, that is as correct as it can be at all temperatures given the relevant idealisations, exists in the literature. The subject routinely hardly rates a mention in textbooks. If we already meet with difficulty when calculating the entropy

of an ideal gas, we might expect bad things farther down the road. And yet in spite of this, statistical mechanics is very successful at predicting and shedding light on some of Nature's very complex behaviour.

Another example of the difficulties underlying the subject can be found in my exposition of blackbody radiation. Every discussion of this subject that I have seen rests on the idea of electromagnetic field nodes at the oven walls. There seems no reason to assume this (especially for non-metallic walls), but if such nodes do exist at the walls, then Planck's law would already have to be a continuous approximation to what is really a discrete function dependent on oven size. Also, the necessary introduction of the concept of emissivity shows that Planck's law applies to an idealised oven only, for emissivity varies with the material of the emitter, and is even a function of wavelength for a single emitter. The key question seems to be just how to define such an idealised oven. I have stressed this approach in these notes.

Parts of these lectures owe much to those given by Graeme Putt and Paul Barker of Auckland University's Physics Department during my own undergraduate physics degree there. My discussions of the course topics here have benefited from conversations with Sanjeev Arulampalam, Shayne Bennetts, Scott Foster, David Griffiths, Alex Kalloniatis, Roland Keir, Jim McCarthy, Jamie Quinton, Andy Rawlinson, Nikita Simakov, Keith Stowe, Alice von Trojan, and Vivienne Wheaton. I wish to thank DSTO and Flinders University for the opportunity to give this lecture course. The clarity of these notes was also greatly improved by the feedback of my class, who were an interested group of students and a delight to teach. There seemed to be no time to really teach everything properly in the few months in which I gave the course, but the students were always willing to listen to my endless asides and my "here he goes again" digressions into semi-relevant mathematics. At least, that's the way it seemed.

References

1. K. Stowe (2007), *An Introduction to Thermodynamics and Statistical Mechanics*, 2nd ed., Cambridge University Press.

2. M.J. Lighthill (1958), *An Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press. See page 32.

3. F. Reif (1965), *Fundamentals of Statistical and Thermal Physics*, Waveland Press, Illinois.

4. D.J. Griffiths (2004), *Introduction to Quantum Mechanics*, 2nd ed., Benjamin Cummings.

5. J.P. Sethna (2006), *Statistical Mechanics: Entropy, Order Parameters and Complexity*, Oxford Master Series in Physics.

Index

Italicised numbers denote where the entry is defined.

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA	1. CAVEAT/PRIVACY MARKING

2. TITLE	3. SECURITY CLASSIFICATION
A Course of Lectures on Statistical Mechanics	Document (U)
	Title (U)
	Abstract (U)

4. AUTHOR	5. CORPORATE AUTHOR
Don Koks	Defence Science and Technology Organisation PO Box 1500 Edinburgh, SA 5111, Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO–GD–0612	AR–014–777	General Document	June 2010

8. FILE NUMBER	9. TASK NUMBER	10. SPONSOR	11. No. OF PAGES	12. No. OF REFS
2010/1001571/1	N/A	DSTO	89	5

13. URL OF ELECTRONIC VERSION	14. RELEASE AUTHORITY
http://www.dsto.defence.gov.au/corporate/ reports/DSTO-GD-0612.pdf	Chief, Electronic Warfare and Radar Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT
Approved for Public Release
OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT
No Limitations

17. CITATION IN OTHER DOCUMENTS
No Limitations

18. DSTO RESEARCH LIBRARY THESAURUS	
Statistical mechanics	Blackbody radiation
Thermodynamics	Temperature
Entropy	Information theory

19. ABSTRACT
This is a set of lectures given by the author in 2009 at Flinders University, Adelaide, comprising one semester of a third-year undergraduate course in physics. The lectures begin with an introduction to the theoretical background of statistical mechanics, and then continue with a mixture of theory and application. Topics covered are those that comprise the standard tool kit for advanced study in the field.